



## Central limit theorems for Diophantine approximants

Downloaded from: <https://research.chalmers.se>, 2023-05-05 03:02 UTC

Citation for the original published paper (version of record):

Björklund, M., Gorodnik, A. (2019). Central limit theorems for Diophantine approximants. *Mathematische Annalen*, 374(3-4): 1371-1437. <http://dx.doi.org/10.1007/s00208-019-01828-1>

N.B. When citing this work, cite the original published paper.



# Central limit theorems for Diophantine approximants

Michael Björklund<sup>1</sup> · Alexander Gorodnik<sup>2</sup>

Received: 21 August 2018 / Revised: 26 February 2019  
© The Author(s) 2019

## Abstract

In this paper we study certain counting functions which represent the numbers of solutions of systems of linear inequalities arising in the theory of Diophantine approximation. We develop a method that allows us to explain the random-like behavior that these functions exhibit and prove a central limit theorem for them. Our approach is based on a quantitative study of higher-order correlations for functions defined on the space of lattices and a novel technique for estimating cumulants of Siegel transforms.

**Mathematics Subject Classification** Primary 11K60; Secondary 60F05 · 37A17

## Contents

1	Introduction and main results	...
2	Estimates on higher-order correlations	...
3	CLT for functions with compact support	...
4	Non-divergence estimates for Siegel transforms	...
5	CLT for smooth Siegel transforms	...
6	CLT for counting functions and the proof of Theorem 1.2	...
	References	...

---

Communicated by Kannan Soundararajan.

---

✉ Michael Björklund  
micbjo@chalmers.se  
Alexander Gorodnik  
alexander.gorodnik@math.uzh.ch

<sup>1</sup> Department of Mathematics, Chalmers, Gothenburg, Sweden

<sup>2</sup> Institut für Mathematik, Universität Zürich, Zürich, Switzerland

# 1 Introduction and main results

## 1.1 Motivation

Many objects which arise in Diophantine Geometry exhibit random-like behavior. For instance, the classical Khinchin theorem in Diophantine approximation can be interpreted as the Borel–Cantelli Property for quasi-independent events, while Schmidt’s quantitative generalization of Khinchin’s Theorem is analogous to the Law of Large Numbers. One might ask whether much deeper probabilistic phenomena also take place.

In this paper, we develop a general framework which allows us to capture certain independence properties which govern the asymptotic behavior of arithmetic counting functions. We expect that the new methods will have a wide range of applications in Diophantine Geometry; here we apply the techniques to study the distribution of a class of counting functions which we now describe.

A basic problem in Diophantine approximation is to find “good” rational approximants of vectors  $\bar{u} = (u_1, \dots, u_m) \in \mathbb{R}^m$ . More precisely, given positive numbers  $w_1, \dots, w_m$ , which we shall assume sum to one, and positive constants  $\vartheta_1, \dots, \vartheta_m$ , we consider the system of inequalities

$$\left| u_j - \frac{p_j}{q} \right| \leq \frac{\vartheta_j}{q^{1+w_j}}, \quad \text{for } j = 1, \dots, m, \quad (1.1)$$

with  $(\bar{p}, q) \in \mathbb{Z}^m \times \mathbb{N}$ . It is well-known that for Lebesgue-almost all  $\bar{u} \in \mathbb{R}^m$ , the system (1.1) has infinitely many solutions  $(\bar{p}, q) \in \mathbb{Z}^m \times \mathbb{N}$ , so it is natural to try to count solutions in bounded regions, which leads us to the counting function

$$\Delta_T(\bar{u}) := |\{(\bar{p}, q) \in \mathbb{Z}^m \times \mathbb{N} : 1 \leq q < T \text{ and (1.1) holds}\}|.$$

Schmidt [15] proved that for Lebesgue-almost all  $\bar{u} \in [0, 1]^m$ ,

$$\Delta_T(\bar{u}) = C_m \log T + O_{\bar{u}, \varepsilon}((\log T)^{1/2+\varepsilon}), \quad \text{for all } \varepsilon > 0, \quad (1.2)$$

where  $C_m := 2^m \vartheta_1 \cdots \vartheta_m$ . One may view this as an analogue of the Law of Large Numbers, the heuristic for this analogy runs along the following lines. First, note that

$$\Delta_T(\bar{u}) \approx \sum_{s=0}^{\lfloor \log T \rfloor} \Delta^{(s)}(\bar{u}),$$

where

$$\Delta^{(s)}(\bar{u}) := |\{(\bar{p}, q) \in \mathbb{Z}^m \times \mathbb{N} : e^s \leq q < e^{s+1} \text{ and (1.1) holds}\}|.$$

If one could prove that the functions  $\Delta^{(s_1)}(\cdot)$  and  $\Delta^{(s_2)}(\cdot)$  were “quasi-independent” random variables on  $[0, 1]^m$ , at least when  $s_1, s_2$ , and  $|s_1 - s_2|$  are sufficiently large,

then (1.2) would follow by some version of the Law of Large Numbers. Moreover, the same heuristic further suggests that, in addition to the Law of Large Numbers, a central limit theorem and perhaps other probabilistic limit laws also hold for  $\Delta_T(\cdot)$ .

In this paper, we put the above heuristic on firm ground. We do so by representing  $\Delta^{(s)}(\cdot)$  as a function on the space of unimodular lattices. It turns out that the “quasi-independence” of the family  $(\Delta^{(s)})$  that we are trying to capture can be translated into the dynamical language of higher-order mixing for a subgroup of linear transformations acting on the space of lattices.

## 1.2 Main results

We are not the first to explore central limit theorems for Diophantine approximants. The one-dimensional case ( $m = 1$ ) has been thoroughly investigated by Leveque [11, 12], Philipp [13], and Fuchs [6], leading to the following result proved by Fuchs [6]: there exists an explicit  $\sigma > 0$  such that the counting function

$$\Delta_T(u) := \left| \left\{ (p, q) \in \mathbb{Z} \times \mathbb{N} : 1 \leq q < T, \quad |u - p/q| < \frac{\vartheta}{q^2 \log(1+q)} \right\} \right|$$

satisfies

$$\left| \left\{ u \in [0, 1] : \frac{\Delta_T(u) - 2\vartheta \log \log T}{(\log \log T \cdot \log \log \log T)^{1/2}} < \xi \right\} \right| \longrightarrow \text{Norm}_\sigma(\xi) \quad (1.3)$$

as  $T \rightarrow \infty$ , where

$$\text{Norm}_\sigma(\xi) := (2\pi\sigma)^{-1/2} \int_{-\infty}^{\xi} e^{-s^2/(2\sigma)} ds$$

denotes the normal distribution with the variance  $\sigma$ .

Central limit theorems in higher dimensions when  $w_1 = \dots = w_m = 1/m$  have recently been studied Dolgopyat et al. [3]. In this paper, using very different techniques, we establish the following CLT for *general* exponents  $w_1, \dots, w_m$ .

**Theorem 1.1** *Let  $m \geq 2$ . Then for every  $\xi \in \mathbb{R}$ ,*

$$\left| \left\{ \bar{u} \in [0, 1]^m : \frac{\Delta_T(\bar{u}) - C_m \log T}{(\log T)^{1/2}} < \xi \right\} \right| \longrightarrow \text{Norm}_{\sigma_m}(\xi) \quad (1.4)$$

as  $T \rightarrow \infty$ , where

$$\sigma_m := 2C_m(2\zeta(m)\zeta(m+1)^{-1} - 1),$$

and  $\zeta$  denotes Riemann’s  $\zeta$ -function.

Our proof of Theorem 1.1, as well as the proof in [3], proceeds by interpreting  $\Delta_T(\cdot)$  as a function on a certain subset  $\mathcal{Y}$  of the space of all unimodular lattices in  $\mathbb{R}^{m+1}$ ,

and then studies how the sequence  $a^s \mathcal{Y}$ , where  $a$  is a fixed linear transformation of  $\mathbb{R}^{m+1}$ , distributes inside this space. However, the arguments in the two papers follow very different routes. The proof in [3] contains a novel refinement of the martingale method (this approach was initiated in this setting by Le Borgne [10]). Here, one crucially uses the fact that when  $w_1 = \dots = w_m = 1/m$ , then the set  $\mathcal{Y}$  is an unstable manifold for the action of  $a$  on the space of lattices. For general weights,  $\mathcal{Y}$  has strictly smaller dimension than the unstable leaves, and it seems challenging to apply martingale approximation techniques. Instead, our method involves a quantitative analysis of higher-order correlations for functions on the space of lattices. We establish an asymptotic formula for correlations of arbitrary orders and use this formula to compute limits of all the moments of  $\Delta_T(\cdot)$  directly. One of the key innovations of our approach is an efficient way of estimating sums of cumulants (alternating sums of moments) developed in our recent work [2].

We also investigate the more general problem of Diophantine approximation for systems of linear forms. The space  $M_{m,n}(\mathbb{R})$  of  $m$  linear forms in  $n$  real variables is parametrized by real  $m \times n$  matrices. Given  $u \in M_{m,n}(\mathbb{R})$ , we consider the family  $(L_u^{(i)})$  of linear forms defined by

$$L_u^{(i)}(x_1, \dots, x_n) = \sum_{j=1}^n u_{ij} x_j, \quad i = 1, \dots, m.$$

Let  $\|\cdot\|$  be a norm on  $\mathbb{R}^n$ . Fix  $\vartheta_1, \dots, \vartheta_m > 0$  and  $w_1, \dots, w_m > 0$  which satisfy

$$w_1 + \dots + w_m = n,$$

and consider the system of Diophantine inequalities

$$\left| p_i + L_u^{(i)}(q_1, \dots, q_n) \right| < \vartheta_i \|\bar{q}\|^{-w_i}, \quad i = 1, \dots, m, \quad (1.5)$$

with  $(\bar{p}, \bar{q}) = (p_1, \dots, p_m, q_1, \dots, q_n) \in \mathbb{Z}^m \times (\mathbb{Z}^n \setminus \{0\})$ . The number of solutions of this system with the norm of the “denominator”  $\bar{q}$  bounded by  $T$  is given by

$$\Delta_T(u) := \left| \{(\bar{p}, \bar{q}) \in \mathbb{Z}^m \times \mathbb{Z}^n : 0 < \|\bar{q}\| < T \text{ and (1.5) holds}\} \right|. \quad (1.6)$$

Our main result in this paper is the following generalization of Theorem 1.1.

**Theorem 1.2** *If  $m \geq 2$ , then for every  $\xi \in \mathbb{R}$ ,*

$$\left| \left\{ u \in M_{m,n}([0, 1]) : \frac{\Delta_T(u) - C_{m,n} \log T}{(\log T)^{1/2}} < \xi \right\} \right| \longrightarrow \text{Norm}_{\sigma_{m,n}}(\xi) \quad (1.7)$$

as  $T \rightarrow \infty$ , where

$$C_{m,n} := C_m \omega_n \quad \text{with } \omega_n := \int_{S^{n-1}} \|\bar{z}\|^{-n} d\bar{z}$$

and

$$\sigma_{m,n} := 2C_{m,n}(2\zeta(m+n-1)\zeta(m+n)^{-1} - 1).$$

The special case  $w_1 = \dots = w_m = n/m$  was proved earlier in [3].

### 1.3 An outline of the proof of Theorem 1.2

We begin by observing that  $\Delta_T(\cdot)$  can be interpreted as a function on the space of lattices in  $\mathbb{R}^{m+n}$ . Given  $u \in \mathbf{M}_{m,n}([0, 1])$ , we define the unimodular lattice  $\Lambda_u$  in  $\mathbb{R}^{m+n}$  by

$$\Lambda_u := \left\{ \left( p_1 + \sum_{j=1}^n u_{1j}q_j, \dots, p_m + \sum_{j=1}^n u_{mj}q_j, \bar{q} \right) : (\bar{p}, \bar{q}) \in \mathbb{Z}^m \times \mathbb{Z}^n \right\}, \quad (1.8)$$

and we see that

$$\Delta_T(u) = |\Lambda_u \cap \Omega_T| + O(1), \quad \text{for } T > 0,$$

where  $\Omega_T$  denotes the domain

$$\Omega_T := \{(\bar{x}, \bar{y}) \in \mathbb{R}^{m+n} : 1 \leq \|\bar{y}\| < T, |x_i| < \vartheta_i \|\bar{y}\|^{-w_i}, i = 1, \dots, m\}. \quad (1.9)$$

The space  $\mathcal{X}$  of unimodular lattices in  $\mathbb{R}^{m+n}$  is naturally a homogeneous space of the group  $\mathrm{SL}_{m+n}(\mathbb{R})$  equipped with the invariant probability measure  $\mu_{\mathcal{X}}$ . The set

$$\mathcal{Y} := \{\Lambda_u : u \in \mathbf{M}_{m,n}([0, 1])\}$$

is a  $mn$ -dimensional torus embedded in  $\mathcal{X}$ , and we equip  $\mathcal{Y}$  with the Haar probability measure  $\mu_{\mathcal{Y}}$ , interpreted as a Borel measure on  $\mathcal{X}$ .

We further observe (see Sect. 6 for more details) that each domain  $\Omega_T$  can be tessellated using a fixed diagonal matrix  $a$  in  $\mathrm{SL}_{m+n}(\mathbb{R})$ , so that for a suitable function  $\hat{\chi} : \mathcal{X} \rightarrow \mathbb{R}$ , we have

$$|\Lambda \cap \Omega_T| \approx \sum_{s=0}^{N-1} \hat{\chi}(a^s \Lambda) \quad \text{for } \Lambda \in \mathcal{X}.$$

Hence we are left with analyzing the distribution of values for the sums  $\sum_{s=0}^N \hat{\chi}(a^s y)$  with  $y \in \mathcal{Y}$ . This will allow us to apply techniques developed in our previous work [2], as well as in [1] (joint with M. Einsiedler). Intuitively, our arguments will be guided by the hope that the observables  $\hat{\chi} \circ a^s$  are “quasi-independent” with respect to  $\mu_{\mathcal{Y}}$ . Due to the discontinuity and unboundedness of the function  $\hat{\chi}$  on  $\mathcal{X}$ , it gets quite technical to formulate this quasi-independence directly. Instead, we shall argue in steps.

We begin in Sect. 2 by establishing quasi-independence for observables of the form  $\phi \circ a^s$ , where  $\phi$  is a smooth and compactly supported function on  $\mathcal{X}$ . This amounts to an asymptotic formula (Corollary 2.4) for the higher-order correlations

$$\int_{\mathcal{Y}} \phi_1(a^{s_1} y) \cdots \phi_r(a^{s_r} y) d\mu_{\mathcal{Y}}(y) \quad \text{with } \phi_1, \dots, \phi_r \in C_c^\infty(\mathcal{X}). \quad (1.10)$$

It will be crucial for our arguments later that the error term in this formula is explicit in terms of the exponents  $s_1, \dots, s_r$  and in (certain norms of) the functions  $\phi_1, \dots, \phi_r$ . In Sect. 3, we use these estimates to prove the central limit theorems for sums of the form

$$F_N(y) := \sum_{s=0}^{N-1} (\phi(a^s y) - \mu_{\mathcal{Y}}(\phi)) \quad \text{with } \phi \in C_c^\infty(\mathcal{X}) \text{ and } y \in \mathcal{Y}.$$

To do this, we use an adaption of the classical Cumulant Method (see Proposition 3.4), which provides bounds on cumulants (alternating sums of moments) given estimates on expressions as in (1.10), at least in certain ranges of the parameters  $(s_1, \dots, s_r)$ . Here we shall exploit the decomposition (3.7) into “separated”/“clustered” tuples. We stress that the cumulant  $\text{Cum}^{(r)}(F_N)$  of order  $r$  can be expressed as a sum of  $O(N^r)$  terms, normalized by  $N^{r/2}$ , so that in order to prove that it vanishes asymptotically, we require more than just square-root cancellation; however, the error term in the asymptotic formula for (1.10) is rather weak. Nonetheless, by using intricate combinatorial cancellations of cumulants, we can establish the required bounds.

In order to extend the method in Sect. 3 to the kind of *unbounded* functions which arise in our subsequent approximation arguments we have to investigate possible escapes of mass for the sequence of tori  $a^s \mathcal{Y}$  inside the space  $\mathcal{X}$ . In Sect. 4, we prove several results in this direction (see e.g. Proposition 4.5), as well as  $L^p$ -bounds (see Propositions 4.6 and 4.8). We stress that the general non-divergence estimates for unipotent flows developed by Kleinbock–Margulis [8] are not sufficient for our purposes, and in particular, the exact value of the exponent in Proposition 4.5 will be crucial for our argument. The proof of the  $L^2$ -norm bound in Proposition 4.8 is especially interesting in this regard since it uncovers that the escape of mass is related to delicate arithmetic questions; our arguments require careful estimates on the number of solutions of certain Diophantine equations.

To make the technical passages in the final steps of the proof of Theorem 1.2 a bit more readable, we shall devote Sect. 5 to central limit theorems for sums of the form

$$\sum_{s=0}^{N-1} \hat{f}(a^s y) \quad \text{for } y \in \mathcal{Y},$$

where  $f$  is a smooth and compactly supported function on  $\mathbb{R}^{m+n}$ , and  $\hat{f}$  denotes the Siegel transform of  $f$  (see Sect. 4.3 for definitions). We stress that even though  $f$  is assumed to be bounded,  $\hat{f}$  is *unbounded* on  $\mathcal{X}$ . To prove the central limit theorems in this setting, we approximate  $\hat{f}$  by compactly supported functions on  $\mathcal{X}$  and then use the estimates from Sect. 3. However, the bounds in these estimates crucially depend

on the order of approximation, so this step requires a delicate analysis of the error terms. The non-divergence results established in Sect. 4 play important role here.

Finally, to prove the central limit theorem for the function  $\hat{\chi}$  (which is the Siegel transform of an indicator function on a nice bounded domain in  $\mathbb{R}^{m+n}$ ), and thus establish Theorem 1.2, we need to approximate  $\chi$  with smooth functions, and show that the arguments in Sect. 5 can be adapted to certain *sequences* of Siegel transforms of smooth and compactly supported functions. This will be done in Sect. 6.

## 2 Estimates on higher-order correlations

Let  $\mathcal{X}$  denote the space of unimodular lattices in  $\mathbb{R}^{m+n}$ . Setting

$$G := \mathrm{SL}_{m+n}(\mathbb{R}) \quad \text{and} \quad \Gamma := \mathrm{SL}_{m+n}(\mathbb{Z}),$$

we may consider the space  $\mathcal{X}$  as a homogeneous space under the linear action of the group  $G$ , so that

$$\mathcal{X} \simeq G/\Gamma.$$

Let  $\mu_{\mathcal{X}}$  denote the  $G$ -invariant probability measure on  $\mathcal{X}$ .

We fix  $m, n \geq 1$  and denote by  $U$  the subgroup

$$U := \left\{ \begin{pmatrix} I_m & u \\ 0 & I_n \end{pmatrix} : u \in M_{m,n}(\mathbb{R}) \right\} < G, \quad (2.1)$$

and set  $\mathcal{Y} := U\mathbb{Z}^{m+n} \subset \mathcal{X}$ . Geometrically,  $\mathcal{Y}$  can be visualized as a  $mn$ -dimensional torus embedded in the spaces of lattices  $\mathcal{X}$ . We denote by  $\mu_{\mathcal{Y}}$  the probability measure on  $\mathcal{Y}$  induced by the Lebesgue probability measure on  $M_{m,n}([0, 1])$ , and we note that  $\mathcal{Y}$  corresponds to the collection of unimodular lattices  $\Lambda_u$ , for  $u \in M_{m,n}([0, 1])$ , introduced earlier in (1.8).

Let us further fix positive numbers  $w_1, \dots, w_{m+n}$  satisfying

$$\sum_{i=1}^m w_i = \sum_{i=m+1}^{m+n} w_i,$$

and denote by  $(a_t)$  the one-parameter semi-group

$$a_t := \mathrm{diag} \left( e^{w_1 t}, \dots, e^{w_m t}, e^{-w_{m+1} t}, \dots, e^{-w_{m+n} t} \right), \quad t > 0. \quad (2.2)$$

The aim of this section is to analyze the asymptotic behavior of  $a_t \mathcal{Y} \subset \mathcal{X}$  as  $t \rightarrow \infty$ , and investigate “decoupling” of correlations of the form

$$\int_{\mathcal{Y}} \phi_1(a_{t_1} y) \dots \phi_r(a_{t_r} y) d\mu_{\mathcal{Y}}(y) \quad \text{for } \phi_1, \dots, \phi_r \in C_c^\infty(\mathcal{X}), \quad (2.3)$$

for “large”  $t_1, \dots, t_r > 0$ . It will be essential for our subsequent argument that the error terms in this “decoupling” are explicit in terms of the parameters  $t_1, \dots, t_r > 0$  and suitable norms of the functions  $\phi_1, \dots, \phi_r$ , which we now introduce.



Every  $Y \in \text{Lie}(G)$  defines a first order differential operator  $\mathcal{D}_Y$  on  $C_c^\infty(\mathcal{X})$  by

$$\mathcal{D}_Y(\phi)(x) := \frac{d}{dt} \phi(\exp(tY)x)|_{t=0}.$$

If we fix an (ordered) basis  $\{Y_1, \dots, Y_r\}$  of  $\text{Lie}(G)$ , then every monomial  $Z = Y_1^{\ell_1} \dots Y_r^{\ell_r}$  defines a differential operator by

$$\mathcal{D}_Z := \mathcal{D}_{Y_1}^{\ell_1} \dots \mathcal{D}_{Y_r}^{\ell_r}, \quad (2.4)$$

of degree  $\deg(Z) = \ell_1 + \dots + \ell_r$ . For  $k \geq 1$  and  $\phi \in C_c^\infty(\mathcal{X})$ , we define the norms

$$\|\phi\|_{L_k^2(\mathcal{X})} := \left( \sum_{\deg(Z) \leq k} \int_{\mathcal{X}} |(\mathcal{D}_Z \phi)(x)|^2 d\mu_{\mathcal{X}}(x) \right)^{1/2}, \quad (2.5)$$

and

$$\|\phi\|_{C^k} := \sum_{\deg(Z) \leq k} \|(\mathcal{D}_Z \phi)(x)\|_{\infty}. \quad (2.6)$$

Note that for every  $g \in G$ ,  $\phi \in C^\infty(\mathcal{X})$  and  $Y \in \text{Lie}(G)$ , we have  $\mathcal{D}_Y(\phi \circ g) = \mathcal{D}_{\text{Ad}(g)Y}(\phi) \circ g$ . This identity readily extends to the universal enveloping algebras  $\mathcal{U}(\text{Lie}(G))$  as well, and thus we also have  $\mathcal{D}_Z(\phi \circ g) = \mathcal{D}_{\text{Ad}(g)Z}(\phi) \circ g$ , for every monomial  $Z$  in  $\{Y_1, \dots, Y_r\}$ , where  $\text{Ad}(g)$  denotes the extension of the  $\text{Ad}(g)$  from  $\text{Lie}(G)$  to  $\mathcal{U}(\text{Lie}(G))$ . Since  $\text{Ad}(g)Z$  can be written as a finite sum of monomials of degrees not exceeding the degree of  $Z$ , we conclude that for every  $k \geq 1$ , there exists a sub-multiplicative function  $g \mapsto C_k(g)$  such that

$$\|\phi \circ g\|_{C^k} \leq C_k(g) \|\phi\|_{C^k}, \quad \text{for all } \phi \in C_c^\infty(\mathcal{X}).$$

In particular, there a constant  $\xi = \xi(m, n, k)$  (which also depends on our fixed choice of weights  $w_1, \dots, w_{m+n}$ ) such that

$$\|\phi \circ a^t\|_{C^k} \ll e^{\xi t} \|\phi\|_{C^k}, \quad \text{for all } t \in \mathbb{R} \text{ and } \phi \in C_c^\infty(\mathcal{X}), \quad (2.7)$$

where the suppressed constants are independent of  $t$  and  $\phi$ .

The starting point of our discussion is a well-known quantitative estimate on correlations of smooth functions on  $\mathcal{X}$ :

**Theorem 2.1** *There exist  $\gamma > 0$  and  $k \geq 1$  such that for all  $\phi_1, \phi_2 \in C_c^\infty(\mathcal{X})$  and  $g \in G$ ,*

$$\begin{aligned} \int_{\mathcal{X}} \phi_1(gx) \phi_2(x) d\mu_{\mathcal{X}}(x) &= \left( \int_{\mathcal{X}} \phi_1 d\mu_{\mathcal{X}} \right) \left( \int_{\mathcal{X}} \phi_2 d\mu_{\mathcal{X}} \right) \\ &\quad + O\left( \|g\|^{-\gamma} \|\phi_1\|_{L_k^2(\mathcal{X})} \|\phi_2\|_{L_k^2(\mathcal{X})} \right). \end{aligned}$$

This theorem has a very long history that we will not attempt to survey here, but only mention that a result of this form can be found, for instance, in [7, Corollary 2.4.4].

From now on we fix  $k \geq 1$  so that Theorem 2.1 holds.

Our goal is to decouple the higher-order correlations in (2.3), but in order to state our results we first need to introduce a family of finer norms on  $C_c^\infty(\mathcal{X})$  than  $(\|\cdot\|_{L^2(\mathcal{X})_k})$ . Let us denote by  $\|\cdot\|_{C^0}$  the uniform norm on  $C_c(\mathcal{X})$ . If we fix a right-invariant Riemannian metric on  $G$ , then it induces a metric  $d$  on  $\mathcal{X} \simeq G/\Gamma$ , which allows us to define the norms

$$\|\phi\|_{Lip} := \sup \left\{ \frac{|\phi(x_1) - \phi(x_2)|}{d(x_1, x_2)} : x_1, x_2 \in \mathcal{X}, x_1 \neq x_2 \right\},$$

and

$$\mathcal{N}_k(\phi) := \max \left\{ \|\phi\|_{C^0}, \|\phi\|_{Lip}, \|\phi\|_{L_k^2(\mathcal{X})} \right\}, \quad (2.8)$$

for  $\phi \in C_c^\infty(\mathcal{X})$ . We shall prove:

**Theorem 2.2** *There exists  $\delta > 0$  such that for every compact  $\Omega \subset U$ ,  $f \in C_c^\infty(U)$  with  $\text{supp}(f) \subset \Omega$ ,  $\phi_1, \dots, \phi_r \in C_c^\infty(\mathcal{X})$ ,  $x_0 \in \mathcal{X}$ , and  $t_1, \dots, t_r > 0$ , we have*

$$\begin{aligned} \int_U f(u) \left( \prod_{i=1}^r \phi_i(a_{t_i} u x_0) \right) du &= \left( \int_U f(u) du \right) \prod_{i=1}^r \left( \int_{\mathcal{X}} \phi_i d\mu_{\mathcal{X}} \right) \\ &\quad + O_{x_0, \Omega, r} \left( e^{-\delta D(t_1, \dots, t_r)} \|f\|_{C^k} \prod_{i=1}^r \mathcal{N}_k(\phi_i) \right), \end{aligned}$$

where

$$D(t_1, \dots, t_r) := \min\{t_i, |t_i - t_j| : 1 \leq i \neq j \leq r\}.$$

**Remark 2.3** The case  $r = 1$  was proved by Kleinbock and Margulis in [9], and our arguments are inspired by theirs. We stress that the constant  $\delta$  in Theorem 2.2 is independent of  $r$ .

We also record the following corollary of Theorem 2.2.

**Corollary 2.4** *There exists  $\delta' > 0$  such that for every  $\phi_0 \in C^\infty(\mathcal{Y})$ ,  $\phi_1, \dots, \phi_r \in C_c^\infty(\mathcal{X})$  and  $t_1, \dots, t_r > 0$ , we have*

$$\begin{aligned} \int_{\mathcal{Y}} \phi_0(y) \left( \prod_{i=1}^r \phi_i(a_{t_i} y) \right) d\mu_{\mathcal{Y}}(y) &= \left( \int_{\mathcal{Y}} \phi_0 d\mu_{\mathcal{Y}} \right) \prod_{i=1}^r \left( \int_{\mathcal{X}} \phi_i d\mu_{\mathcal{X}} \right) \\ &\quad + O_r \left( e^{-\delta' D(t_1, \dots, t_r)} \|\phi_0\|_{C^k} \prod_{i=1}^r \mathcal{N}_k(\phi_i) \right). \end{aligned}$$

**Proof of Corollary 2.4 (assuming Theorem 2.2)** Let  $x_0$  denote the identity coset in  $\mathcal{X} \cong G/\Gamma$ , which corresponds to the standard lattice  $\mathbb{Z}^{m+n}$ , and recall that

$$\mathcal{Y} = Ux_0 \simeq U/(U \cap \Gamma).$$

Let  $\tilde{\phi}_0 \in C^\infty(U)$  denote the lift of the function  $\phi_0$  to  $U$ , and  $\chi$  the characteristic function of the subset

$$U_0 := \left\{ \begin{pmatrix} I_m & u \\ 0 & I_n \end{pmatrix} : u \in M_{m,n}([0, 1]) \right\}.$$

Given  $\varepsilon > 0$ , let  $\chi_\varepsilon \in C_c^\infty(U)$  be a smooth approximation of  $\chi$  with uniformly bounded support which satisfy

$$\chi \leq \chi_\varepsilon \leq 1, \quad \|\chi - \chi_\varepsilon\|_{L^1(U)} \ll \varepsilon, \quad \|\chi_\varepsilon\|_{C^k} \ll \varepsilon^{-k}.$$

We observe that if  $f_\varepsilon := \tilde{\phi}_0 \chi_\varepsilon$  and  $f_0 := \tilde{\phi}_0 \chi$ , then

$$\|f_0 - f_\varepsilon\|_{L^1(U)} \ll \varepsilon \|\phi_0\|_{C^0}$$

and

$$\|f_\varepsilon\|_{C^k} \ll \|\tilde{\phi}_0\|_{C^k} \|\chi_\varepsilon\|_{C^k} \ll \varepsilon^{-k} \|\phi_0\|_{C^k},$$

which implies that

$$\begin{aligned} \int_{\mathcal{Y}} \phi_0(y) \left( \prod_{i=1}^r \phi_i(a_{t_i} y) \right) d\mu_{\mathcal{Y}}(y) &= \int_U f_0(u) \left( \prod_{i=1}^r \phi_i(a_{t_i} u x_0) \right) du \\ &= \int_U f_\varepsilon(u) \left( \prod_{i=1}^r \phi_i(a_{t_i} u x_0) \right) du \\ &\quad + O \left( \varepsilon \prod_{i=0}^r \|\phi_i\|_{C^0} \right), \end{aligned}$$

and

$$\int_{\mathcal{Y}} \phi_0 d\mu_{\mathcal{Y}} = \int_U f_0(u) du = \int_U f_\varepsilon(u) du + O(\varepsilon \|\phi_0\|_{C^0}).$$

Therefore, Theorem 2.2 implies that

$$\int_{\mathcal{Y}} \phi_0(y) \left( \prod_{i=1}^r \phi_i(a_{t_i} y) \right) d\mu_{\mathcal{Y}}(y) = \left( \int_U f_\varepsilon(u) du \right) \prod_{i=1}^r \left( \int_{\mathcal{X}} \phi_i d\mu_{\mathcal{X}} \right)$$

$$\begin{aligned}
 & + O_r \left( \varepsilon \prod_{i=0}^r \|\phi_i\|_{C^0} + e^{-\delta D(t_1, \dots, t_r)} \|f_\varepsilon\|_{C^k} \prod_{i=1}^r \mathcal{N}_k(\phi_i) \right) \\
 & = \left( \int_{\mathcal{Y}} \phi_0 d\mu_{\mathcal{Y}} \right) \prod_{i=1}^r \left( \int_{\mathcal{X}} \phi_i d\mu_{\mathcal{X}} \right) \\
 & + O_r \left( \left( \varepsilon + \varepsilon^{-k} e^{-\delta D(t_1, \dots, t_r)} \right) \|\phi_0\|_{C^k} \prod_{i=1}^r \mathcal{N}_k(\phi_i) \right).
 \end{aligned}$$

The corollary (with  $\delta' = \delta/(k+1)$ ) follows by choosing  $\varepsilon = e^{-\delta D(t_1, \dots, t_r)/(k+1)}$ .  $\square$

## 2.1 Preliminary results

We recall that  $d$  is a distance on  $\mathcal{X} \cong G/\Gamma$  induced from a right-invariant Riemannian metric on  $G$ . We denote by  $B_G(\rho)$  the ball of radius  $\rho$  centered at the identity in  $G$ . For a point  $x \in \mathcal{X}$ , we let  $\iota(x)$  denote the *injectivity radius* at  $x$ , that is to say, the supremum over  $\rho > 0$  such that the map  $B_G(\rho) \rightarrow B_G(\rho)x : g \mapsto gx$  is injective.

Given  $\varepsilon > 0$ , let

$$\mathcal{K}_\varepsilon = \{\Lambda \in \mathcal{X} : \|v\| \geq \varepsilon, \text{ for all } v \in \Lambda \setminus \{0\}\}. \quad (2.9)$$

By Mahler's Compactness Criterion,  $\mathcal{K}_\varepsilon$  is a compact subset of  $\mathcal{X}$ . Furthermore, using reduction theory, one can show:

**Proposition 2.5** [9, Prop. 3.5]  $\iota(x) \gg \varepsilon^{m+n}$  for any  $x \in \mathcal{K}_\varepsilon$ .

An important role in our argument will be played by the one-parameter semi-group

$$b_t := \text{diag} \left( e^{t/m}, \dots, e^{t/m}, e^{-t/n}, \dots, e^{-t/n} \right), \quad t > 0, \quad (2.10)$$

which coincides with the semi-group  $(a_t)$  as defined in (2.2) with the special choice of exponents

$$w_1 = \dots = w_m = \frac{1}{m} \quad \text{and} \quad w_{m+1} = \dots = w_{m+n} = \frac{1}{n}.$$

The submanifold  $\mathcal{Y} \subset \mathcal{X}$  is an unstable manifold for the flow  $(b_t)$  which makes the analysis of the asymptotic behavior of  $b_t \mathcal{Y}$  significantly easier than that of  $a_t \mathcal{Y}$  for general parameters. Using Theorem 2.1, Kleinbock and Margulis proved in [7] a quantitative equidistribution result for the family  $b_t \mathcal{Y}$  as  $t \rightarrow \infty$ , we shall use a version of this result from their later work [9].

**Theorem 2.6** [9, Th. 2.3] *There exist  $\rho_0 > 0$  and  $c, \gamma > 0$  such that for every  $\rho \in (0, \rho_0)$ ,  $f \in C_c^\infty(U)$  satisfying  $\text{supp}(f) \subset B_G(\rho)$ ,  $x \in \mathcal{X}$  with  $\iota(x) > 2\rho$ ,  $\phi \in C_c^\infty(\mathcal{X})$ , and  $t \geq 0$ ,*

$$\int_U f(u)\phi(b_t u x) du = \left( \int_U f(u) du \right) \left( \int_{\mathcal{X}} \phi d\mu_{\mathcal{X}} \right) + O\left( \rho \|f\|_{L^1(U)} \|\phi\|_{\text{Lip}} + \rho^{-c} e^{-\gamma t} \|f\|_{C^k} \|\phi\|_{L_k^2(\mathcal{X})} \right).$$

**Remark 2.7** Although the dependence on  $\phi$  is not stated in [9, Theorem 2.3], the estimate is explicit in the proof. Indeed, in [9, Section 2, p. 390], the authors show under the assumptions above that

$$\left| \int_U f(u)\phi(b_t u x) du - \langle \tilde{f}, b_{-t}\phi \rangle_{L^2(\mathcal{X})} \right| \leq \rho \|\phi\|_{\text{Lip}} \|f\|_{L^1(U)},$$

where  $\tilde{f}$  is an (explicit) smooth function on  $\mathcal{X}$  with compact support constructed from  $f$ . Theorem 2.6 now follows from the decay of matrix coefficients in Theorem 2.1.

We will prove Theorem 2.2 through successive uses of Theorem 2.6. In order to make things more transparent, it will be convenient to embed the flow  $(a_t)$  as defined in (2.2) in a multi-parameter flow as follows. For  $\bar{s} = (s_1, \dots, s_{m+n}) \in \mathbb{R}^{m+n}$ , we set

$$a(\bar{s}) := \text{diag} \left( e^{s_1}, \dots, e^{s_m}, e^{-s_{m+1}}, \dots, e^{-s_{m+n}} \right). \quad (2.11)$$

We denote by  $S^+$  the cone in  $\mathbb{R}^{m+n}$  consisting of those  $\bar{s} = (s_1, \dots, s_{m+n})$  which satisfy

$$s_1, \dots, s_{m+n} > 0 \quad \text{and} \quad \sum_{i=1}^m s_i = \sum_{i=m+1}^{m+n} s_i.$$

For  $\bar{s} = (s_1, \dots, s_{m+n}) \in S^+$ , we set

$$[\bar{s}] := \min(s_1, \dots, s_{m+n}),$$

and, with  $\bar{s}_t := (w_1 t, \dots, w_{m+n} t)$ , we see that  $a_t = a(\bar{s}_t)$ .

In addition to Theorem 2.6, we shall also need the following quantitative non-divergence estimate for unipotent flows established by Kleinbock and Margulis in [9].

**Theorem 2.8** [9, Cor. 3.4] *There exists  $\theta = \theta(m, n) > 0$  such that for every compact  $L \subset \mathcal{X}$  and a Euclidean ball  $B \subset U$  centered at the identity, there exists  $T_0 > 0$  such that for every  $\varepsilon \in (0, 1)$ ,  $x \in L$ , and  $\bar{s} \in S^+$  satisfying  $[\bar{s}] \geq T_0$ , one has*

$$|\{u \in B : a(\bar{s})ux \notin \mathcal{K}_\varepsilon\}| \ll \varepsilon^\theta |B|.$$

## 2.2 Proof of Theorem 2.2

Let us fix  $r \geq 1$  and a  $r$ -tuple  $(t_1, \dots, t_r)$ . Upon re-labeling, we may assume that  $t_1 \leq \dots \leq t_r$ , so that

$$D := D(t_1, \dots, t_r) = \min\{t_1, t_2 - t_1, \dots, t_r - t_{r-1}\}. \quad (2.12)$$

The next lemma provides an additional parameter  $\bar{s} \in S^+$ , which depends on the  $r$ -tuple  $(t_1, \dots, t_r)$ . This parameter will be used throughout the proof of Theorem 2.2, and we stress that the accompanying constants  $c_1$ ,  $c_2$  and  $c_3$  are independent of  $r$  and the  $r$ -tuple  $(t_1, \dots, t_r)$ .

**Lemma 2.9** *There exist  $c_1, c_2, c_3 > 0$  such that given any  $t_r > t_{r-1} > 0$ , there exists  $\bar{s} \in S^+$  satisfying:*

- (i)  $\lfloor \bar{s} \rfloor \geq c_1(t_r - t_{r-1})$ ,
- (ii)  $\lfloor \bar{s} - \bar{s}_{t_{r-1}} \rfloor \geq c_2(t_r - t_{r-1})$ ,
- (iii)  $\bar{s}_{t_r} - \bar{s} = (\frac{z}{m}, \dots, \frac{z}{m}, \frac{z}{n}, \dots, \frac{z}{n})$  for some  $z \geq c_3 \min(t_{r-1}, t_r - t_{r-1})$ .

**Proof** We start the proof by defining  $\bar{s}$  by the formula in (iii), where the parameter  $z$  will be chosen later, that is to say, we set

$$\bar{s} = \left( w_1 t_r - \frac{z}{m}, \dots, w_m t_r - \frac{z}{m}, w_{m+1} t_r - \frac{z}{n}, \dots, w_{m+n} t_r - \frac{z}{n} \right).$$

Then (i) holds provided that

$$A_1 t_r - \frac{z}{m} \geq c_1(t_r - t_{r-1}) \quad \text{and} \quad A_2 t_r - \frac{z}{n} \geq c_1(t_r - t_{r-1}),$$

where  $A_1 := \min(w_i : 1 \leq i \leq m)$  and  $A_2 := \min(w_i : m+1 \leq i \leq m+n)$ , so if we set  $c_1 = \min(A_1, A_2)$ , then (i) holds when

$$z \leq c_1 \min(m, n) t_{r-1}. \quad (2.13)$$

To arrange (ii), we observe that

$$\begin{aligned} \bar{s} - \bar{s}_{t_{r-1}} &= \left( w_1(t_r - t_{r-1}) - \frac{z}{m}, \dots, w_m(t_r - t_{r-1}) - \frac{z}{m}, \right. \\ &\quad \left. \times w_{m+1}(t_r - t_{r-1}) - \frac{z}{n}, \dots, w_{m+n}(t_r - t_{r-1}) - \frac{z}{n} \right), \end{aligned}$$

and thus (ii) holds provided that

$$A_1(t_r - t_{r-1}) - \frac{z}{m} \geq c_2(t_r - t_{r-1}) \quad \text{and} \quad A_2(t_r - t_{r-1}) - \frac{z}{n} \geq c_2(t_r - t_{r-1}).$$

If we let  $c_2 = \min(A_1, A_2)/2$ , then (ii) holds when

$$z \leq c_2 \min(m, n)(t_r - t_{r-1}). \quad (2.14)$$

So far we have arranged so that (i) and (ii) hold provided that  $z$  satisfies (2.13) and (2.14). Let  $c_3 = \min(c_1, c_2) \min(m, n)$ , and note that if we pick  $z = c_3 \min(t_{r-1}, t_r - t_{r-1})$ , then (i), (ii) and (iii) are all satisfied.  $\square$

Let us now continue with the proof of Theorem 2.2. With the parameter  $\bar{s}$  provided by Lemma 2.9 above, we have

$$\lfloor \bar{s} \rfloor \geq c_1 D, \quad (2.15)$$

$$\lfloor \bar{s} - \bar{s}_{t_i} \rfloor \geq c_2 D \quad \text{for all } i = 1, \dots, r-1, \quad (2.16)$$

$$a(\bar{s}_{t_r} - \bar{s}) = b_z \quad \text{for some } z \geq c_3 D, \quad (2.17)$$

where  $b_z$  is defined as in (2.10) and  $D$  as in (2.12). Throughout the rest of the proof we fix a compact set  $\Omega \subset U$ . Our aim now is to estimate integrals of the form

$$I_r := \int_U f(u) \left( \prod_{i=1}^r \phi_i(a_{t_i} u x_0) \right) du,$$

where  $f$  ranges over  $C_c^\infty(U)$  with  $\text{supp}(f) \subset \Omega$ . Our proof will proceed by induction over  $r$ , the case  $r = 0$  being trivial.

Before we can start the induction, we need some notation. Let  $\rho_0$  and  $k$  be as in Theorem 2.6, and pick for  $0 < \rho < \rho_0$ , a non-negative  $\omega_\rho \in C_c^\infty(\mathcal{X})$  such that

$$\text{supp}(\omega_\rho) \subset B_G(\rho), \quad \|\omega_\rho\|_{C^k} \ll \rho^{-\sigma}, \quad \int_U \omega_\rho(v) dv = 1, \quad (2.18)$$

for some fixed  $\sigma = \sigma(m, n, k) > 0$ . The integral  $I_r$  can now be rewritten as follows:

$$\begin{aligned} I_r &= I_r \cdot \left( \int_U \omega_\rho(v) dv \right) = \int_{U \times U} f(u) \omega_\rho(v) \left( \prod_{i=1}^r \phi_i(a_{t_i} u x_0) \right) dudv \\ &= \int_{U \times U} f(a(-\bar{s})va(\bar{s})u) \omega_\rho(v) \left( \prod_{i=1}^r \phi_i(a_{t_i} a(-\bar{s})va(\bar{s})ux_0) \right) dudv. \end{aligned}$$

If we set

$$f_{\bar{s},u}(v) := f(a(-\bar{s})va(\bar{s})u) \omega_\rho(v) \quad \text{and} \quad x_{\bar{s},u} := a(\bar{s})ux_0,$$

then

$$I_r = \int_U \left( \int_U f_{\bar{s},u}(v) \prod_{i=1}^r \phi_i(a(\bar{s}_{t_i} - \bar{s})v x_{\bar{s},u}) dv \right) du.$$

We observe that if  $f_{\bar{s},u}(v) \neq 0$ , then

$$v \in \text{supp}(\omega_\rho) \subset B_G(\rho) \quad \text{and} \quad a(-\bar{s})va(\bar{s})u \in \text{supp}(f),$$

so that

$$u \in a(-\bar{s})v^{-1}a(\bar{s})\text{supp}(f) \subset a(-\bar{s})\text{supp}(\omega_\rho)^{-1}a(\bar{s})\Omega.$$

Since  $\bar{s} \in S^+$ , the linear map  $v \mapsto a(-\bar{s})va(\bar{s})$ , for  $v \in U \cong \mathbb{R}^{mn}$ , is *non-expanding*, and thus we can conclude that there exists a fixed Euclidean ball  $B$  in  $U$  (depending only on  $\Omega$  and  $\rho_0$ ), and centered at the identity, such that if  $f_{\bar{s},u}(v) \neq 0$ , then  $u \in B$ . This implies that the integral  $I_r$  can be written as

$$I_r = \int_B \left( \int_U f_{\bar{s},u}(v) \prod_{i=1}^r \phi_i(a(\bar{s}_{t_i} - \bar{s})vx_{\bar{s},u}) dv \right) du, \quad (2.19)$$

and

$$\|f_{\bar{s},u}\|_{C^k} \ll \|f\|_{C^k} \|\omega_\rho\|_{C^k} \ll \rho^{-\sigma} \|f\|_{C^k}. \quad (2.20)$$

Furthermore,

$$\begin{aligned} \int_U \|f_{\bar{s},u}\|_{L^1(U)} du &= \int_{U \times U} |f(a(-\bar{s})va(\bar{s})u)\omega_\rho(v)| dv du \\ &= \left( \int_U |f(u)| du \right) \left( \int_U \omega_\rho(v) dv \right) = \|f\|_{L^1(U)}. \end{aligned} \quad (2.21)$$

We decompose the integral  $I_r$  in (2.19) as

$$I_r = I'_r(\varepsilon) + I''_r(\varepsilon),$$

where  $I'_r(\varepsilon)$  is the integral over

$$B_\varepsilon := \{u \in B : x_{\bar{s},u} \notin \mathcal{K}_\varepsilon\},$$

and  $I''_r(\varepsilon)$  is the integral over  $B \setminus B_\varepsilon$ .

To estimate  $I'_r(\varepsilon)$ , we first recall that  $\bar{s} \geq c_1 D$  by (2.15), so if  $D \geq T_0/c_1$ , where  $T_0$  is as in Theorem 2.8 applied to  $\mathcal{L} = \mathcal{K}_\varepsilon$  and  $B$ , then the same theorem implies that there exists  $\theta > 0$  such that

$$|B_\varepsilon| \ll \varepsilon^\theta |B| \quad (2.22)$$

for every  $\varepsilon \in (0, 1)$ , and thus

$$I'_r(\varepsilon) \ll_\Omega \varepsilon^\theta |B| \|f\|_{C^0} \left( \int_U \omega_\rho(v) dv \right) \prod_{i=1}^r \|\phi_i\|_{C^0} \ll_\Omega \varepsilon^\theta \|f\|_{C^0} \prod_{i=1}^r \|\phi_i\|_{C^0}. \quad (2.23)$$

Let us now turn to the problem of estimating  $I''_r(\varepsilon)$ . Since the Riemannian distance  $d$  on  $G$ , restricted to  $U$ , and the Euclidean distance on  $U$  are equivalent on a small open identity neighborhood, we see that

$$d(a(-\bar{t})va(\bar{t}), e) \ll e^{-[\bar{t}]} d(v, e).$$



for  $v \in B_G(\rho_0)$  and any  $\bar{t} \in S^+$ . Hence, using (2.16), we obtain that for all  $i = 1, \dots, r-1$ ,

$$\phi_i(a(\bar{s}_{t_i} - \bar{s})vx_{\bar{s},u}) = \phi_i(a(\bar{s}_{t_i} - \bar{s})x_{\bar{s},u}) + O\left(e^{-c_2D} \|\phi_i\|_{Lip}\right),$$

and thus, for all  $v \in B_G(\rho_0)$ ,

$$\prod_{i=1}^{r-1} \phi_i(a(\bar{s}_{t_i} - \bar{s})vx_{\bar{s},u}) = \prod_{i=1}^{r-1} \phi_i(a(\bar{s}_{t_i} - \bar{s})x_{\bar{s},u}) + O_r\left(e^{-c_2D} \prod_{i=1}^{r-1} \|\phi_i\|_{Lip}^*\right), \quad (2.24)$$

where  $\|\phi\|_{Lip}^* := \max(\|\phi\|_{C^0}, \|\phi\|_{Lip})$ . This leads to the estimate

$$\begin{aligned} I_r''(\varepsilon) &= \int_{B \setminus B_\varepsilon} \prod_{i=1}^{r-1} \phi_i(a(\bar{s}_{t_i} - \bar{s})x_{\bar{s},u}) \left( \int_U f_{\bar{s},u}(v) \phi_r(a(\bar{s}_{t_r} - \bar{s})vx_{\bar{s},u}) dv \right) du \\ &\quad + O_r\left(e^{-c_2D} \prod_{i=1}^{r-1} \|\phi_i\|_{Lip}^* \left( \int_{B \setminus B_\varepsilon} \|f_{\bar{s},u}\|_{L^1(U)} du \right) \|\phi_r\|_{C^0}\right). \end{aligned}$$

Hence, using (2.21), we obtain that

$$\begin{aligned} I_r''(\varepsilon) &= \int_{B \setminus B_\varepsilon} \prod_{i=1}^{r-1} \phi_i(a(\bar{s}_{t_i} - \bar{s})x_{\bar{s},u}) \left( \int_U f_{\bar{s},u}(v) \phi_r(a(\bar{s}_{t_r} - \bar{s})vx_{\bar{s},u}) dv \right) du \\ &\quad + O_r\left(e^{-c_2D} \|f\|_{L^1(U)} \prod_{i=1}^r \|\phi_i\|_{Lip}^*\right). \end{aligned}$$

Since by (2.17),

$$\int_U f_{\bar{s},u}(v) \phi_r(a(\bar{s}_{t_r} - \bar{s})vx_{\bar{s},u}) dv = \int_U f_{\bar{s},u}(v) \phi_r(b_z vx_{\bar{s},u}) dv,$$

we apply Theorem 2.6 to estimate this integral. We recall that  $\text{supp}(f_{\bar{s},u}) \subset B_G(\rho)$ . For  $u \in B \setminus B_\varepsilon$ , we have  $x_{\bar{s},u} \in \mathcal{K}_\varepsilon$ , so that  $\iota(x_{\bar{s},u}) \gg \varepsilon^{m+n}$  by Proposition 2.5. In particular, we may take

$$\varepsilon \gg \rho^{1/(m+n)}$$

to arrange that  $\iota(x_{\bar{s},u}) > 2\rho$ . Hence, by applying Theorem 2.6, we deduce that there exist  $c, \gamma > 0$  such that

$$\begin{aligned} \int_U f_{\bar{s},u}(v) \phi_r(b_z vx_{\bar{s},u}) dv &= \left( \int_U f_{\bar{s},u}(v) dv \right) \left( \int_{\mathcal{X}} \phi_r d\mu_{\mathcal{X}} \right) \\ &\quad + O\left(\rho \|f_{\bar{s},u}\|_{L^1(U)} \|\phi_r\|_{Lip} + \rho^{-c} e^{-\gamma z} \|f_{\bar{s},u}\|_{C^k} \|\phi_r\|_{L_k^2(\mathcal{X})}\right), \end{aligned}$$

for all  $u \in B \setminus B_\varepsilon$ . Using (2.20) and (2.21) and  $z \geq c_3 D$ , we deduce that

$$\begin{aligned} I_r''(\varepsilon) &= \left( \int_{B \setminus B_\varepsilon} \prod_{i=1}^{r-1} \phi_i(a(\bar{s}_{t_i} - \bar{s})x_{\bar{s},u}) \left( \int_U f_{\bar{s},u}(v) dv \right) du \right) \left( \int_{\mathcal{X}} \phi_r d\mu_{\mathcal{X}} \right) \\ &\quad + O_{r,B} \left( \prod_{i=1}^{r-1} \|\phi_i\|_{C^0} \left( \rho \|f\|_{L^1(U)} \|\phi_r\|_{Lip} + \rho^{-(c+\sigma)} e^{-\gamma c_3 D} \|f\|_{C^k} \|\phi_r\|_{L_k^2(\mathcal{X})} \right) \right) \\ &\quad + O_r \left( e^{-c_2 D} \|f\|_{L^1(U)} \prod_{i=1}^r \|\phi_i\|_{Lip}^* \right). \end{aligned}$$

Since  $\|f\|_{L^1(U)} \ll_{\Omega} \|f\|_{C^k}$ ,

$$\begin{aligned} I_r''(\varepsilon) &= \left( \int_{B \setminus B_\varepsilon} \prod_{i=1}^{r-1} \phi_i(a(\bar{s}_{t_i} - \bar{s})x_{\bar{s},u}) \left( \int_U f_{\bar{s},u}(v) dv \right) du \right) \left( \int_{\mathcal{X}} \phi_r d\mu_{\mathcal{X}} \right) \\ &\quad + O_{r,\Omega} \left( \left( \rho + \rho^{-(c+\sigma)} e^{-\gamma c_3 D} + e^{-c_2 D} \right) \|f\|_{C^k} \prod_{i=1}^r \mathcal{N}_k(\phi_i) \right). \quad (2.25) \end{aligned}$$

Applying (2.24) one more time (in the backward direction), we get

$$\begin{aligned} &\int_{B \setminus B_\varepsilon} \prod_{i=1}^{r-1} \phi_i(a(\bar{s}_{t_i} - \bar{s})x_{\bar{s},u}) \left( \int_U f_{\bar{s},u}(v) dv \right) du \\ &= \int_{B \setminus B_\varepsilon} \left( \int_U f_{\bar{s},u}(v) \prod_{i=1}^{r-1} \phi_i(a(\bar{s}_{t_i} - \bar{s})vx_{\bar{s},u}) dv \right) du \\ &\quad + O_r \left( e^{-c_2 D} \left( \int_B \|f_{\bar{s},u}\|_{L^1(U)} du \right) \prod_{i=1}^{r-1} \|\phi_i\|_{Lip}^* \right). \end{aligned}$$

It follows from (2.22) that

$$\begin{aligned} &\int_{B \setminus B_\varepsilon} \left( \int_U f_{\bar{s},u}(v) \prod_{i=1}^{r-1} \phi_i(a(\bar{s}_{t_i} - \bar{s})vx_{\bar{s},u}) dv \right) du \\ &= \int_B \left( \int_U f_{\bar{s},u}(v) \prod_{i=1}^{r-1} \phi_i(a(\bar{s}_{t_i} - \bar{s})vx_{\bar{s},u}) dv \right) du \\ &\quad + O_r \left( \varepsilon^\theta \left( \int_B \|f_{\bar{s},u}\|_{L^1(U)} du \right) \prod_{i=1}^{r-1} \|\phi_i\|_{C^0} \right), \end{aligned}$$

where we recognize the first term as  $I_{r-1}$ . Using (2.19) and (2.21), we now conclude that

$$\begin{aligned} & \int_{B \setminus B_\varepsilon} \prod_{i=1}^{r-1} \phi_i(a(\bar{s}_{i_i} - \bar{s})x_{\bar{s},u}) \left( \int_U f_{\bar{s},u}(v) dv \right) du \\ &= I_{r-1} + O_r \left( \left( \varepsilon^\theta + e^{-c_2 D} \right) \|f\|_{L^1(U)} \prod_{i=1}^{r-1} \|\phi_i\|_{Lip}^* \right). \end{aligned}$$

Hence, combining this estimate with (2.25), we deduce that

$$\begin{aligned} I_r''(\varepsilon) &= I_{r-1} \left( \int_{\mathcal{X}} \phi_r d\mu_{\mathcal{X}} \right) \\ &+ O_{r,\Omega} \left( \left( \varepsilon^\theta + \rho + \rho^{-(c+\sigma)} e^{-\gamma c_3 D} + e^{-c_2 D} \right) \|f\|_{C^k} \prod_{i=1}^r \mathcal{N}_k(\phi_i) \right), \end{aligned}$$

and thus, in view of (2.23),

$$\begin{aligned} I_r &= I_r'(\varepsilon) + I_r''(\varepsilon) = I_{r-1} \left( \int_{\mathcal{X}} \phi_r d\mu_{\mathcal{X}} \right) \\ &+ O_{r,\Omega} \left( \left( \varepsilon^\theta + \rho + \rho^{-(c+\sigma)} e^{-\gamma c_3 D} + e^{-c_2 D} \right) \|f\|_{C^k} \prod_{i=1}^r \mathcal{N}_k(\phi_i) \right). \end{aligned}$$

This estimate holds whenever  $\rho < \rho_0$  and  $\varepsilon \gg \rho^{1/(m+n)}$ . Taking  $\rho = e^{-c_4 D}$  for sufficiently small  $c_4 > 0$  and  $\varepsilon \ll \rho^{1/(m+n)}$ , we conclude that there exists  $\delta > 0$  such that for all sufficiently large  $D$ ,

$$I_r = I_{r-1} \left( \int_{\mathcal{X}} \phi_r d\mu_{\mathcal{X}} \right) + O_{r,\Omega} \left( e^{-\delta D} \|f\|_{C^k} \prod_{i=1}^r \mathcal{N}_k(\phi_i) \right). \quad (2.26)$$

The exponent  $\delta$  depends on the constants  $c_2$  and  $c_3$  given by Lemma 2.9 and the parameters  $\theta, c, \sigma, \gamma$ . In particular,  $\delta$  is independent of  $r$ . By possibly enlarging the implicit constants we can ensure that the estimate (2.26) also holds for all  $r$ -tuples  $(t_1, \dots, t_r)$ , and not just the ones with sufficiently large  $D(t_1, \dots, t_r)$ . By iterating the estimate (2.26), using that  $I_0$  is a constant, the proof of Theorem 2.2 is finished.

### 3 CLT for functions with compact support

Let  $a = \text{diag}(a_1, \dots, a_{m+n})$  be a diagonal linear map of  $\mathbb{R}^{m+n}$  with

$$a_1, \dots, a_m > 1, \quad 0 < a_{m+1}, \dots, a_{m+n} < 1, \quad \text{and} \quad a_1 \cdots a_{m+n} = 1.$$

The map  $a$  defines a continuous self-map of the space  $\mathcal{X}$ , which preserves  $\mu_{\mathcal{X}}$ . We recall that the torus  $\mathcal{Y} = U\mathbb{Z}^{m+n} \subset \mathcal{X}$  is equipped with the probability measure  $\mu_{\mathcal{Y}}$ . In this section, we shall prove a central limit theorem for the averages

$$F_N := \frac{1}{\sqrt{N}} \sum_{s=0}^{N-1} (\phi \circ a^s - \mu_{\mathcal{Y}}(\phi \circ a^s)), \quad (3.1)$$

restricted to  $\mathcal{Y}$ , for a fixed function  $\phi \in C_c^\infty(\mathcal{X})$ . We stress that this result is not needed in the proof of Theorem 1.2, but we nevertheless include it here because we feel that its proof might be instructive before entering the proof of the similar, but far more technical, Theorem 6.1.

**Theorem 3.1** *For every  $\xi \in \mathbb{R}$ ,*

$$\mu_{\mathcal{Y}}(\{y \in \mathcal{Y} : F_N(y) < \xi\}) \rightarrow \text{Norm}_{\sigma_\phi}(\xi)$$

as  $N \rightarrow \infty$ , where

$$\sigma_\phi^2 := \sum_{s=-\infty}^{\infty} \left( \int_{\mathcal{X}} (\phi \circ a^s) \phi \, d\mu_{\mathcal{X}} - \mu_{\mathcal{X}}(\phi)^2 \right).$$

**Remark 3.2** It follows from Theorem 2.1 that the variance  $\sigma_\phi$  is finite.

Our main tool in the proof of Theorem 3.1 will be the estimates on higher-order correlations established in Sect. 2. To make notations less heavy, we shall use a simplified version of Corollary 2.4 stated in terms of  $C^k$ -norms (note that  $\mathcal{N}_k \ll \|\cdot\|_{C^k}$ ):

**Corollary 3.3** *There exists  $\delta > 0$  such that for every  $\phi_0, \dots, \phi_r \in C_c^\infty(\mathcal{X})$  and  $t_1, \dots, t_r > 0$ , we have*

$$\begin{aligned} \int_{\mathcal{Y}} \phi_0(y) \left( \prod_{i=1}^r \phi_i(a_{t_i} y) \right) d\mu_{\mathcal{Y}}(y) &= \left( \int_{\mathcal{Y}} \phi_0 d\mu_{\mathcal{Y}} \right) \prod_{i=1}^r \left( \int_{\mathcal{X}} \phi_i d\mu_{\mathcal{X}} \right) \\ &\quad + O_r \left( e^{-\delta D(t_1, \dots, t_r)} \prod_{i=0}^r \|\phi_i\|_{C^k} \right). \end{aligned}$$

### 3.1 The method of cumulants

Let  $(\mathcal{X}, \mu)$  be a probability space. Given bounded measurable functions  $\phi_1, \dots, \phi_r$  on  $\mathcal{X}$ , we define their *joint cumulant* as

$$\text{Cum}_\mu^{(r)}(\phi_1, \dots, \phi_r) = \sum_{\mathcal{P}} (-1)^{|\mathcal{P}|-1} (|\mathcal{P}| - 1)! \prod_{I \in \mathcal{P}} \int_{\mathcal{X}} \left( \prod_{i \in I} \phi_i \right) d\mu,$$

where the sum is taken over all partitions  $\mathcal{P}$  of the set  $\{1, \dots, r\}$ . When it is clear from the context, we skip the subscript  $\mu$ . For a bounded measurable function  $\phi$  on  $\mathcal{X}$ , we also set

$$\text{Cum}^{(r)}(\phi) = \text{Cum}^{(r)}(\phi, \dots, \phi).$$

We shall use the following classical CLT-criterion (see, for instance, [5]).

**Proposition 3.4** *Let  $(F_T)$  be a sequence of real-valued bounded measurable functions such that*

$$\int_{\mathcal{X}} F_T d\mu = 0 \quad \text{and} \quad \sigma^2 := \lim_{T \rightarrow \infty} \int_{\mathcal{X}} F_T^2 d\mu < \infty \quad (3.2)$$

and

$$\lim_{T \rightarrow \infty} \text{Cum}^{(r)}(F_T) = 0, \quad \text{for all } r \geq 3. \quad (3.3)$$

Then for every  $\xi \in \mathbb{R}$ ,

$$\mu(\{F_T < \xi\}) \rightarrow \text{Norm}_{\sigma}(\xi) \quad \text{as } T \rightarrow \infty.$$

Since all the moments of a random variable can be expressed in terms of its cumulants, this criterion is equivalent to the more widely known “Method of Moments”. However, the cumulants have curious, and very convenient, cancellation properties that will play an important role in our proof of Theorem 3.1.

For a partition  $\mathcal{Q}$  of  $\{1, \dots, r\}$ , we define the *conditional joint cumulant* with respect to  $\mathcal{Q}$  as

$$\text{Cum}_{\mu}^{(r)}(\phi_1, \dots, \phi_r | \mathcal{Q}) = \sum_{\mathcal{P}} (-1)^{|\mathcal{P}|-1} (|\mathcal{P}| - 1)! \prod_{I \in \mathcal{P}} \prod_{J \in \mathcal{Q}} \int_{\mathcal{X}} \left( \prod_{i \in I \cap J} \phi_i \right) d\mu,$$

In what follows, we shall make frequent use of the following proposition.

**Proposition 3.5** [2, Prop. 8.1] *For any partition  $\mathcal{Q}$  with  $|\mathcal{Q}| \geq 2$ ,*

$$\text{Cum}_{\mu}^{(r)}(\phi_1, \dots, \phi_r | \mathcal{Q}) = 0, \quad (3.4)$$

for all  $\phi_1, \dots, \phi_r \in L^{\infty}(\mathcal{X}, \mu)$ .

### 3.2 Estimating cumulants

Fix  $\phi \in C_c^{\infty}(\mathcal{X})$ . It will be convenient to write  $\psi_s(y) := \phi(a^s y) - \mu_{\mathcal{Y}}(\phi \circ a^s)$ , so that

$$F_N := \frac{1}{\sqrt{N}} \sum_{s=0}^{N-1} \psi_s \quad \text{and} \quad \int_{\mathcal{Y}} \psi_s d\mu_{\mathcal{Y}} = 0.$$

In this section, we shall estimate cumulants of the form

$$\text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(F_N) = \frac{1}{N^{r/2}} \sum_{s_1, \dots, s_r=0}^{N-1} \text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(\psi_{s_1}, \dots, \psi_{s_r}) \quad (3.5)$$

for  $r \geq 3$ . Since we shall later need to apply these estimate in cases when the function  $\phi$  is allowed to vary with  $N$ , it will be important to keep track of the dependence on  $\phi$  in our estimates.

We shall decompose (3.5) into sub-sums where the parameters  $s_1, \dots, s_r$  are either “separated” or “clustered”, and it will also be important to control their sizes. For this purpose, it will be convenient to consider the set  $\{0, \dots, N-1\}^r$  as a subset of  $\mathbb{R}_+^{r+1}$  via the embedding  $(s_1, \dots, s_r) \mapsto (0, s_1, \dots, s_r)$ . Following the ideas developed in the paper [2], we define for non-empty subsets  $I$  and  $J$  of  $\{0, \dots, r\}$  and  $\bar{s} = (s_0, \dots, s_r) \in \mathbb{R}_+^{r+1}$ ,

$$\rho^I(\bar{s}) := \max\{|s_i - s_j| : i, j \in I\} \quad \text{and} \quad \rho_{I,J}(\bar{s}) := \min\{|s_i - s_j| : i \in I, j \in J\},$$

and if  $\mathcal{Q}$  is a partition of  $\{0, \dots, r\}$ , we set

$$\rho^{\mathcal{Q}}(\bar{s}) := \max\{\rho^I(\bar{s}) : I \in \mathcal{Q}\} \quad \text{and} \quad \rho_{\mathcal{Q}}(\bar{s}) := \min\{\rho_{I,J}(\bar{s}) : I \neq J, I, J \in \mathcal{Q}\}.$$

For  $0 \leq \alpha < \beta$ , we define

$$\Delta_{\mathcal{Q}}(\alpha, \beta) := \{\bar{s} \in \mathbb{R}_+^{r+1} : \rho^{\mathcal{Q}}(\bar{s}) \leq \alpha, \text{ and } \rho_{\mathcal{Q}}(\bar{s}) > \beta\}$$

and

$$\Delta(\alpha) := \{\bar{s} \in \mathbb{R}_+^{r+1} : |s_i - s_j| \leq \alpha \text{ for all } i, j\}.$$

The following decomposition of  $\mathbb{R}_+^{r+1}$  was established in our paper [2, Prop. 6.2]: given

$$0 = \alpha_0 < \beta_1 < \alpha_1 = (3+r)\beta_1 < \beta_2 < \dots < \beta_r < \alpha_r = (3+r)\beta_r < \beta_{r+1}, \quad (3.6)$$

we have

$$\mathbb{R}_+^{r+1} = \Delta(\beta_{r+1}) \cup \left( \bigcup_{j=0}^r \bigcup_{|\mathcal{Q}| \geq 2} \Delta_{\mathcal{Q}}(\alpha_j, \beta_{j+1}) \right), \quad (3.7)$$

where the union is taken over the partitions  $\mathcal{Q}$  of  $\{0, \dots, r\}$  with  $|\mathcal{Q}| \geq 2$ . Intersecting with  $\{0, 1, \dots, N-1\}^r$ , we see that

$$\{0, \dots, N-1\}^r = \Omega(\beta_{r+1}; N) \cup \left( \bigcup_{j=0}^r \bigcup_{|\mathcal{Q}| \geq 2} \Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; N) \right), \quad (3.8)$$

for all  $N \geq 2$ , where

$$\Omega(\beta_{r+1}; N) := \{0, \dots, N-1\}^r \cap \Delta(\beta_{r+1}),$$

$$\Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; N) := \{0, \dots, N-1\}^r \cap \Delta_{\mathcal{Q}}(\alpha_j, \beta_{j+1}).$$

In order to estimate the cumulant (3.5), we shall separately estimate the sums over  $\Omega(\beta_{r+1}; N)$  and  $\Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; N)$ , the exact choices of the sequences  $(\alpha_j)$  and  $(\beta_j)$  will be fixed at the very end of our argument.

### 3.2.1 Case 0: Summing over $\Omega(\beta_{r+1}; N)$

In this case,  $s_i \leq \beta_{r+1}$  for all  $i$ , and thus

$$|\Omega(\beta_{r+1}; N)| \leq (\beta_{r+1} + 1)^r.$$

Hence,

$$\sum_{(s_1, \dots, s_r) \in \Omega(\beta_{r+1}; N)} \text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(\psi_{s_1}, \dots, \psi_{s_r}) \ll (\beta_{r+1} + 1)^r \|\phi\|_{C^0}^r, \quad (3.9)$$

where the implied constants may depend on  $r$ , but we shall henceforth omit this subscript to simplify notations.

### 3.2.2 Case 1: Summing over $\Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; N)$ with $\mathcal{Q} = \{\{0\}, \{1, \dots, r\}\}$

In this case, we have

$$|s_{i_1} - s_{i_2}| \leq \alpha_j \quad \text{for all } i_1, i_2,$$

so that it follows that

$$|\Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; N)| \ll N(1 + \alpha_j)^{r-1}.$$

Hence,

$$\frac{1}{N^{r/2}} \sum_{(s_1, \dots, s_r) \in \Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; N)} |\text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(\psi_{s_1}, \dots, \psi_{s_r})| \ll N^{1-r/2} \alpha_j^{r-1} \|\phi\|_{C^0}^r. \quad (3.10)$$

### 3.2.3 Case 2: Summing over $\Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; N)$ with $|\mathcal{Q}| \geq 2$ and $\mathcal{Q} \neq \{\{0\}, \{1, \dots, r\}\}$

In this case, the partition  $\mathcal{Q}$  defines a non-trivial partition  $\mathcal{Q}' = \{I_0, \dots, I_\ell\}$  of  $\{1, \dots, r\}$  such that for all  $(s_1, \dots, s_r) \in \Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; N)$ , we have

$$|s_{i_1} - s_{i_2}| \leq \alpha_j \quad \text{if } i_1 \sim_{\mathcal{Q}'} i_2 \quad \text{and} \quad |s_{i_1} - s_{i_2}| > \beta_{j+1} \quad \text{if } i_1 \not\sim_{\mathcal{Q}'} i_2, \quad (3.11)$$

and

$$s_i \leq \alpha_j \text{ for all } i \in I_0, \quad \text{and} \quad s_i > \beta_{j+1} \text{ for all } i \notin I_0.$$

In particular,

$$D(s_{i_1}, \dots, s_{i_\ell}) \geq \beta_{j+1}. \quad (3.12)$$

Let  $I$  be an arbitrary subset of  $\{1, \dots, r\}$ . In what follows, we shall show a precise version of the “decoupling”:

$$\int_{\mathcal{Y}} \left( \prod_{i \in I} \psi_{s_i} \right) d\mu_{\mathcal{Y}} \approx \prod_{h=0}^{\ell} \left( \int_{\mathcal{Y}} \left( \prod_{i \in I \cap I_h} \psi_{s_i} \right) d\mu_{\mathcal{Y}} \right), \quad (3.13)$$

where we henceforth shall use the convention that the product is equal to one when  $I \cap I_h = \emptyset$ .

Let us estimate the left hand side of (3.13). We begin by setting

$$\Phi_0 := \prod_{i \in I \cap I_0} \psi_{s_i}.$$

By (2.7), there exists  $\xi = \xi(m, n, k) > 0$  such that

$$\|\Phi_0\|_{C^k} \ll \prod_{i \in I \cap I_0} \|\phi \circ a^{s_i} - \mu_{\mathcal{Y}}(\phi \circ a^{s_i})\|_{C^k} \ll e^{|I \cap I_0| \xi \alpha_j} \|\phi\|_{C^k}^{|I \cap I_0|}. \quad (3.14)$$

To prove (3.13), we expand  $\psi_{s_i} = \phi \circ a^{s_i} - \mu_{\mathcal{Y}}(\phi \circ a^{s_i})$  for  $i \in I \setminus I_0$  and get

$$\begin{aligned} \int_{\mathcal{Y}} \left( \prod_{i \in I} \psi_{s_i} \right) d\mu_{\mathcal{Y}} &= \sum_{J \subset I \setminus I_0} (-1)^{|I \setminus (J \cup I_0)|} \\ &\cdot \left( \int_{\mathcal{Y}} \Phi_0 \left( \prod_{i \in J} \phi \circ a^{s_i} \right) d\mu_{\mathcal{Y}} \right) \prod_{i \in I \setminus (J \cup I_0)} \left( \int_{\mathcal{Y}} (\phi \circ a^{s_i}) d\mu_{\mathcal{Y}} \right). \end{aligned} \quad (3.15)$$

We recall that when  $i \notin I_0$ , we have  $s_i \geq \beta_{j+1}$ , and thus it follows from Corollary 3.3 with  $r = 1$  that

$$\int_{\mathcal{Y}} (\phi \circ a^{s_i}) d\mu_{\mathcal{Y}} = \mu_{\mathcal{X}}(\phi) + O(e^{-\delta \beta_{j+1}} \|\phi\|_{C^k}), \quad \text{with } i \notin I_0. \quad (3.16)$$

To estimate the other integrals in (3.15), we also apply Corollary 3.3. Let us first fix a subset  $J \subset I \setminus I_0$  and for each  $1 \leq h \leq l$ , we pick  $i_h \in I_h$ , and set

$$\Phi_h := \prod_{i \in J \cap I_h} \phi \circ a^{s_i - s_{i_h}}.$$



Then

$$\int_{\mathcal{Y}} \Phi_0 \left( \prod_{i \in J} \phi \circ a^{s_i} \right) d\mu_{\mathcal{Y}} = \int_{\mathcal{Y}} \Phi_0 \left( \prod_{h=1}^{\ell} \Phi_h \circ a^{s_{i_h}} \right) d\mu_{\mathcal{Y}}.$$

We note that for  $i \in I_h$ , we have  $|s_i - s_{i_h}| \leq \alpha_j$ , and thus, by (2.7), there exists  $\xi = \xi(m, n, k) > 0$  such that

$$\|\Phi_h\|_{C^k} \ll \prod_{i \in J \cap I_h} \|\phi \circ a^{s_i - s_{i_h}}\|_{C^k} \ll e^{|J \cap I_h| \xi \alpha_j} \|\phi\|_{C^k}^{|J \cap I_h|}. \quad (3.17)$$

Using (3.12), Corollary 3.3 implies that

$$\begin{aligned} \int_{\mathcal{Y}} \Phi_0 \left( \prod_{h=1}^{\ell} \Phi_h \circ a^{s_{i_h}} \right) d\mu_{\mathcal{Y}} &= \left( \int_{\mathcal{Y}} \Phi_0 d\mu_{\mathcal{Y}} \right) \prod_{h=1}^{\ell} \left( \int_{\mathcal{X}} \Phi_h d\mu_{\mathcal{X}} \right) \\ &\quad + O \left( e^{-\delta\beta_{j+1}} \prod_{h=0}^{\ell} \|\Phi_h\|_{C^k} \right). \end{aligned}$$

Using (3.14) and (3.17) and invariance of the measure  $\mu_{\mathcal{X}}$ , we deduce that

$$\begin{aligned} \int_{\mathcal{Y}} \Phi_0 \left( \prod_{h=1}^{\ell} \Phi_h \circ a^{s_{i_h}} \right) d\mu_{\mathcal{Y}} &= \left( \int_{\mathcal{Y}} \Phi_0 d\mu_{\mathcal{Y}} \right) \prod_{h=1}^{\ell} \left( \int_{\mathcal{X}} \left( \prod_{i \in J \cap I_h} \phi \circ a^{s_i} \right) d\mu_{\mathcal{X}} \right) \\ &\quad + O \left( e^{-(\delta\beta_{j+1} - r\xi\alpha_j)} \|\phi\|_{C^k}^{|(I \cap I_0) \cup J|} \right). \end{aligned}$$

Hence, we conclude that

$$\begin{aligned} \int_{\mathcal{Y}} \Phi_0 \left( \prod_{i \in J} \phi \circ a^{s_i} \right) d\mu_{\mathcal{Y}} &= \left( \int_{\mathcal{Y}} \Phi_0 d\mu_{\mathcal{Y}} \right) \prod_{h=1}^{\ell} \left( \int_{\mathcal{X}} \left( \prod_{i \in J \cap I_h} \phi \circ a^{s_i} \right) d\mu_{\mathcal{X}} \right) \\ &\quad + O \left( e^{-(\delta\beta_{j+1} - r\xi\alpha_j)} \|\phi\|_{C^k}^{|(I \cap I_0) \cup J|} \right). \end{aligned} \quad (3.18)$$

We shall choose the parameters  $\alpha_j$  and  $\beta_{j+1}$  so that

$$\delta\beta_{j+1} - r\xi\alpha_j > 0. \quad (3.19)$$

Substituting (3.16) and (3.18) in (3.15), we deduce that

$$\begin{aligned} &\int_{\mathcal{Y}} \left( \prod_{i \in I} \psi_{s_i} \right) d\mu_{\mathcal{Y}} \\ &= \sum_{J \subset I \setminus I_0} (-1)^{|I \setminus (J \cup I_0)|} \left( \int_{\mathcal{Y}} \Phi_0 d\mu_{\mathcal{Y}} \right) \prod_{h=1}^{\ell} \left( \int_{\mathcal{X}} \left( \prod_{i \in J \cap I_h} \phi \circ a^{s_i} \right) d\mu_{\mathcal{X}} \right) \end{aligned}$$

$$\times \mu_{\mathcal{X}}(\phi)^{|I \setminus (J \cup I_0)|} + O\left(e^{-(\delta\beta_{j+1} - r\xi\alpha_j)} \|\phi\|_{C^k}^{|J|}\right). \quad (3.20)$$

Next, we estimate the right hand side of (3.13). Let us fix  $1 \leq h \leq l$  and for a subset  $J \subset I \cap I_h$ , we define

$$\Phi_J := \prod_{i \in J} \phi \circ a^{s_i - s_{i_h}}.$$

As in (3.17), for some  $\xi > 0$ ,

$$\|\Phi_J\|_{C^k} \ll \prod_{i \in J} \|\phi \circ a^{s_i - s_{i_h}}\|_{C^k} \ll e^{|J|\xi\alpha_j} \|\phi\|_{C^k}^{|J|}.$$

Applying Corollary 3.3 to the function  $\Phi_J$  and using that  $s_{i_h} > \beta_{j+1}$ , we get

$$\begin{aligned} \int_{\mathcal{Y}} \left( \prod_{i \in J} \phi \circ a^{s_i} \right) d\mu_{\mathcal{Y}} &= \int_{\mathcal{Y}} (\Phi_J \circ a^{s_{i_h}}) d\mu_{\mathcal{Y}} \\ &= \int_{\mathcal{X}} \Phi_J d\mu_{\mathcal{X}} + O\left(e^{-\delta\beta_{j+1}} \|\Phi_J\|_{C^k}\right) \\ &= \int_{\mathcal{X}} \left( \prod_{i \in J} \phi \circ a^{s_i} \right) d\mu_{\mathcal{X}} + O\left(e^{-\delta\beta_{j+1}} e^{r\xi\alpha_j} \|\phi\|_{C^k}^{|J|}\right), \end{aligned} \quad (3.21)$$

where we have used  $a$ -invariance of  $\mu_{\mathcal{X}}$ . Combining (3.16) and (3.21), we deduce that

$$\begin{aligned} \int_{\mathcal{Y}} \left( \prod_{i \in I \cap I_h} \psi_{s_i} \right) d\mu_{\mathcal{Y}} &= \sum_{J \subset I \cap I_h} (-1)^{|(I \cap I_h) \setminus J|} \left( \int_{\mathcal{X}} \left( \prod_{i \in J} \phi \circ a^{s_i} \right) d\mu_{\mathcal{X}} \right) \mu_{\mathcal{X}}(\phi)^{|(I \cap I_h) \setminus J|} \\ &\quad + O\left(e^{-\delta\beta_{j+1}} e^{r\xi\alpha_j} \|\phi\|_{C^k}^{|I \cap I_h|}\right) \\ &= \int_{\mathcal{X}} \prod_{i \in I \cap I_h} (\phi \circ a^{s_i} - \mu_{\mathcal{X}}(\phi)) d\mu_{\mathcal{X}} \\ &\quad + O\left(e^{-(\delta\beta_{j+1} - r\xi\alpha_j)} \|\phi\|_{C^k}^{|I \cap I_h|}\right), \end{aligned} \quad (3.22)$$

which implies

$$\begin{aligned} &\prod_{h=0}^{\ell} \left( \int_{\mathcal{Y}} \left( \prod_{i \in I \cap I_h} \psi_{s_i} \right) d\mu_{\mathcal{Y}} \right) \\ &= \left( \int_{\mathcal{Y}} \Phi_0 d\mu_{\mathcal{Y}} \right) \prod_{h=1}^{\ell} \left( \int_{\mathcal{X}} \prod_{i \in I \cap I_h} (\phi \circ a^{s_i} - \mu_{\mathcal{X}}(\phi)) d\mu_{\mathcal{X}} \right) \\ &\quad + O\left(e^{-(\delta\beta_{j+1} - r\xi\alpha_j)} \|\phi\|_{C^k}^{\ell}\right). \end{aligned}$$

Furthermore, multiplying out the products over  $I \cap I_h$ , we get

$$\begin{aligned} & \prod_{h=0}^{\ell} \left( \int_{\mathcal{Y}} \left( \prod_{i \in I \cap I_h} \psi_{s_i} \right) d\mu_{\mathcal{Y}} \right) \\ &= \left( \int_{\mathcal{Y}} \Phi_0 d\mu_{\mathcal{Y}} \right) \sum_{J \subset I \setminus I_0} (-1)^{|I \setminus (I_0 \cup J)|} \prod_{h=1}^{\ell} \left( \int_{\mathcal{X}} \prod_{i \in I_h \cap J} \phi \circ \alpha^{s_i} d\mu_{\mathcal{X}} \right) \mu_{\mathcal{X}}(\phi)^{|I \setminus (I_0 \cup J)|} \\ &+ O \left( e^{-(\delta\beta_{j+1} - r\xi\alpha_j)} \|\phi\|_{C^k}^{|I|} \right). \end{aligned} \quad (3.23)$$

Comparing (3.20) and (3.23), we finally conclude that

$$\begin{aligned} \int_{\mathcal{Y}} \left( \prod_{i \in I} \psi_{s_i} \right) d\mu_{\mathcal{Y}} &= \prod_{h=0}^{\ell} \left( \int_{\mathcal{Y}} \left( \prod_{i \in I \cap I_h} \psi_{s_i} \right) d\mu_{\mathcal{Y}} \right) \\ &+ O \left( e^{-(\delta\beta_{j+1} - r\xi\alpha_j)} \|\phi\|_{C^k}^{|I|} \right) \end{aligned}$$

when  $(s_1, \dots, s_r) \in \Omega_Q(\alpha_j, \beta_{j+1}; N)$ . This establishes (3.13) with an explicit error term. This estimate implies that for the partition  $\mathcal{Q}' = \{I_0, \dots, I_{\ell}\}$ ,

$$\text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(\psi_{s_1}, \dots, \psi_{s_r}) = \text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(\psi_{s_1}, \dots, \psi_{s_r} | \mathcal{Q}') + O \left( e^{-(\delta\beta_{j+1} - r\xi\alpha_j)} \|\phi\|_{C^k}^r \right)$$

By Proposition 3.5,

$$\text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(\psi_{s_1}, \dots, \psi_{s_r} | \mathcal{Q}') = 0,$$

so it follows that for all  $(s_1, \dots, s_r) \in \Omega_Q(\alpha_j, \beta_{j+1}; N)$ ,

$$\left| \text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(\psi_{s_1}, \dots, \psi_{s_r}) \right| \ll e^{-(\delta\beta_{j+1} - r\xi\alpha_j)} \|\phi\|_{C^k}^r. \quad (3.24)$$

### 3.2.4 Final estimates on the cumulants

Let us now return to (3.5). Upon decomposing this sum into the regions discussed above, and applying the estimates (3.9), (3.10) and (3.24) to respective region, we get the bound

$$\begin{aligned} \left| \text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(F_N) \right| &\ll \left( (\beta_{r+1} + 1)^r N^{-r/2} + \left( \max_j \alpha_j^{r-1} \right) N^{1-r/2} \right) \|\phi\|_{C^0}^r \\ &+ N^{r/2} \left( \max_j e^{-(\delta\beta_{j+1} - r\xi\alpha_j)} \right) \|\phi\|_{C^k}^r. \end{aligned} \quad (3.25)$$

This estimate holds provided that (3.6) and (3.19) hold, namely when

$$\alpha_j = (3 + r)\beta_j < \beta_{j+1} \quad \text{and} \quad \delta\beta_{j+1} - r\xi\alpha_j > 0 \quad \text{for } j = 1, \dots, r. \quad (3.26)$$

Given any  $\gamma > 0$ , we define the parameters  $\beta_j$  inductively by the formula

$$\beta_1 = \gamma \quad \text{and} \quad \beta_{j+1} = \max(\gamma + (3+r)\beta_j, \gamma + \delta^{-1}r(3+r)\xi\beta_j). \quad (3.27)$$

It easily follows by induction that  $\beta_{r+1} \ll_r \gamma$ , and we deduce from (3.25) that

$$|\text{Cum}_{\mu_Y}^{(r)}(F_N)| \ll ((\gamma + 1)^r N^{-r/2} + \gamma^{r-1} N^{1-r/2}) \|\phi\|_{C^0}^r + N^{r/2} e^{-\delta\gamma} \|\phi\|_{C^k}^r.$$

Taking  $\gamma = (r/\delta) \log N$ , we conclude that when  $r \geq 3$ ,

$$\text{Cum}_{\mu_Y}^{(r)}(F_N) \rightarrow 0 \quad \text{as } N \rightarrow \infty. \quad (3.28)$$

### 3.3 Estimating the variance

In this section, we wish to compute the limit

$$\|F_N\|_{L^2(\mathcal{Y})}^2 = \frac{1}{N} \sum_{s_1=0}^{N-1} \sum_{s_2=0}^{N-1} \int_{\mathcal{Y}} \psi_{s_1} \psi_{s_2} d\mu_Y.$$

Setting  $s_1 = s + t$  and  $s_2 = t$ , we rewrite the above sums as

$$\|F_N\|_{L^2(\mathcal{Y})}^2 = \Theta_N(0) + 2 \sum_{s=1}^{N-1} \Theta_N(s), \quad (3.29)$$

where

$$\Theta_N(s) := \frac{1}{N} \sum_{t=1}^{N-1-s} \int_{\mathcal{Y}} \psi_{s+t} \psi_t d\mu_Y.$$

Since  $\psi_t = \phi \circ a^t - \mu_Y(\phi \circ a^t)$ ,

$$\int_{\mathcal{Y}} \psi_{s+t} \psi_t d\mu_Y = \int_{\mathcal{Y}} (\phi \circ a^{s+t})(\phi \circ a^t) d\mu_Y - \mu_Y(\phi \circ a^{s+t}) \mu_Y(\phi \circ a^t). \quad (3.30)$$

It follows from Corollary 3.3 that for fixed  $s$  and as  $t \rightarrow \infty$ ,

$$\int_{\mathcal{Y}} (\phi \circ a^{s+t})(\phi \circ a^t) d\mu_Y \rightarrow \int_{\mathcal{X}} (\phi \circ a^s) \phi d\mu_{\mathcal{X}}, \quad \text{and} \quad \mu_Y(\phi \circ a^t) \rightarrow \mu_{\mathcal{X}}(\phi).$$

We conclude that

$$\int_{\mathcal{Y}} \psi_{s+t} \psi_t d\mu_Y \rightarrow \Theta_{\infty}(s) := \int_{\mathcal{X}} (\phi \circ a^s) \phi d\mu_{\mathcal{X}} - \mu_{\mathcal{X}}(\phi)^2,$$

as  $t \rightarrow \infty$ , and for fixed  $s$ ,

$$\Theta_N(s) \rightarrow \Theta_\infty(s) \quad \text{as } N \rightarrow \infty.$$

If one carelessly interchange limits above, one *expects* that as  $N \rightarrow \infty$ ,

$$\|F_N\|_{L^2(\mathcal{Y})}^2 \rightarrow \Theta_\infty(0) + 2 \sum_{s=1}^{\infty} \Theta_\infty(s) = \sum_{s=-\infty}^{\infty} \Theta_\infty(s). \quad (3.31)$$

To prove this limit rigorously, we need to say a bit more to ensure, say, dominated convergence.

It follows from Corollary 3.3 that

$$\begin{aligned} \int_{\mathcal{Y}} (\phi \circ a^{s+t})(\phi \circ a^t) d\mu_{\mathcal{Y}} &= \mu_{\mathcal{X}}(\phi)^2 + O\left(e^{-\delta \min(s,t)} \|\phi\|_{C^k}^2\right), \\ \int_{\mathcal{Y}} (\phi \circ a^{s+t}) d\mu_{\mathcal{Y}} &= \mu_{\mathcal{X}}(\phi) + O\left(e^{-\delta(s+t)} \|\phi\|_{C^k}\right), \\ \int_{\mathcal{Y}} (\phi \circ a^t) d\mu_{\mathcal{Y}} &= \mu_{\mathcal{X}}(\phi) + O\left(e^{-\delta t} \|\phi\|_{C^k}\right). \end{aligned}$$

and thus, in combination with (3.30),

$$\left| \int_{\mathcal{Y}} \psi_{s+t} \psi_t d\mu_{\mathcal{Y}} \right| \ll e^{-\delta \min(s,t)} \|\phi\|_{C^k}^2. \quad (3.32)$$

This integral can also be estimated in a different way. If we set  $\phi_t = \phi \circ a^t$ , then we deduce from Corollary 3.3 that

$$\begin{aligned} \int_{\mathcal{Y}} (\phi \circ a^{s+t})(\phi \circ a^t) d\mu_{\mathcal{Y}} &= \int_{\mathcal{Y}} (\phi_t \circ a^s) \phi_t d\mu_{\mathcal{Y}} \\ &= \mu_{\mathcal{Y}}(\phi_t) \mu_{\mathcal{X}}(\phi_t) + O\left(e^{-\delta s} \|\phi_t\|_{C^k}^2\right), \end{aligned}$$

and

$$\mu_{\mathcal{Y}}(\phi \circ a^{s+t}) = \mu_{\mathcal{Y}}(\phi_t \circ a^s) = \mu_{\mathcal{X}}(\phi_t) + O\left(e^{-\delta s} \|\phi_t\|_{C^k}\right).$$

By (2.7), there exists  $\xi = \xi(m, n, k) > 0$ , such that

$$\|\phi_t\|_{C^k} \ll e^{\xi t} \|\phi\|_{C^k} \quad \text{and} \quad |\mu_{\mathcal{Y}}(\phi_t)| \leq \|\phi\|_{C^k},$$

and thus

$$\left| \int_{\mathcal{Y}} \psi_{s+t} \psi_t d\mu_{\mathcal{Y}} \right| \ll e^{-\delta s} e^{\xi t} \|\phi\|_{C^k}^2. \quad (3.33)$$

Let us now combine (3.32) and (3.33): When  $t \leq \delta/(2\xi)s$ , we use (3.33), and when  $t \geq \delta/(2\xi)s$ , we use (3.32). If we set  $\delta' = \min(\delta/2, \delta^2/(2\xi)) > 0$ , then

$$\left| \int_{\mathcal{Y}} \psi_{s+t} \psi_t d\mu_{\mathcal{Y}} \right| \ll e^{-\delta' s} \|\phi\|_{C^k}^2$$

for all  $s \geq 0$ , whence

$$|\Theta_N(s)| \leq \frac{1}{N} \sum_{t=1}^{N-1-s} \left| \int_{\mathcal{Y}} \psi_{s+t} \psi_t d\mu_{\mathcal{Y}} \right| \ll e^{-\delta' s} \|\phi\|_{C^k}^2$$

uniformly in  $N$ . Hence, the Dominated Convergence Theorem applied to (3.29) yields (3.31).

### 3.4 Proof of Theorem 3.1

In this subsection we shall check that the conditions of Proposition 3.4 hold for the sequence  $(F_N)$  defined in (3.1). First, by construction, it is easy to check that

$$\int_{\mathcal{Y}} F_T d\mu_{\mathcal{Y}} = 0,$$

and by (3.31),

$$\int_{\mathcal{Y}} F_N^2 d\mu_{\mathcal{Y}} \rightarrow \sum_{s=-\infty}^{\infty} \Theta_{\infty}(s) \quad \text{as } N \rightarrow \infty.$$

Furthermore, by (3.28),

$$\text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(F_N) \rightarrow 0 \quad \text{as } N \rightarrow \infty,$$

for every  $r \geq 3$ , which finishes the proof.

## 4 Non-divergence estimates for Siegel transforms

### 4.1 Siegel transforms

We recall that the space  $\mathcal{X}$  of unimodular lattices in  $\mathbb{R}^{m+n}$  can be identified with the quotient space  $G/\Gamma$ , where  $G = \text{SL}_{m+n}(\mathbb{R})$  and  $\Gamma = \text{SL}_{m+n}(\mathbb{Z})$ , which is endowed with the  $G$ -invariant probability measures  $\mu_{\mathcal{X}}$ . We denote by  $m_G$  a bi- $G$ -invariant Radon measure on  $G$ . Given a bounded measurable function  $f : \mathbb{R}^{m+n} \rightarrow \mathbb{R}$  with compact support, we define its *Siegel transform*  $\hat{f} : \mathcal{X} \rightarrow \mathbb{R}$  by

$$\hat{f}(\Lambda) := \sum_{z \in \Lambda \setminus \{0\}} f(z) \quad \text{for } \Lambda \in \mathcal{X}.$$

We stress that  $\hat{f}$  is unbounded on  $\mathcal{X}$ , its growth is controlled by an explicit function  $\alpha$  which we now introduce. Given a lattice  $\Lambda$  in  $\mathbb{R}^{m+n}$ , we say that a subspace  $V$  of  $\mathbb{R}^{m+n}$  is  $\Lambda$ -rational if the intersection  $V \cap \Lambda$  is a lattice in  $V$ . If  $V$  is  $\Lambda$ -rational, we denote by  $d_\Lambda(V)$  the volume of  $V/(V \cap \Lambda)$ , and define

$$\alpha(\Lambda) := \sup \left\{ d_\Lambda(V)^{-1} : V \text{ is a } \Lambda\text{-rational subspace of } \mathbb{R}^{m+n} \right\}.$$

It follows from the Mahler Compactness Criterion that  $\alpha$  is a proper function on  $\mathcal{X}$ .

**Proposition 4.1** [16, Lem. 2] *If  $f : \mathbb{R}^{m+n} \rightarrow \mathbb{R}$  is a bounded function with compact support, then*

$$|\hat{f}(\Lambda)| \ll_{\text{supp}(f)} \|f\|_{C^0} \alpha(\Lambda) \quad \text{for all } \Lambda \in \mathcal{X}.$$

Using reduction theory, it is not hard to derive the following integrability of  $\alpha$ , which is well-known in Geometry of Numbers (see e.g. [4, Lemma 3.10]).

**Proposition 4.2**  $\alpha \in L^p(\mathcal{X})$  for  $1 \leq p < m+n$ . In particular,

$$\mu_{\mathcal{X}}(\{\alpha \geq L\}) \ll_p L^{-p} \quad \text{for all } p < m+n.$$

In what follows,  $d\bar{z}$  denotes the volume element on  $\mathbb{R}^{m+n}$  which assigns volume one to the unit cube. In our arguments below, we will make heavy use of the following two integral formulas:

**Proposition 4.3** (Siegel Mean Value Theorem; [17]) *If  $f : \mathbb{R}^{m+n} \rightarrow \mathbb{R}$  is a bounded Riemann integrable function with compact support, then*

$$\int_{\mathcal{X}} \hat{f} d\mu_{\mathcal{X}} = \int_{\mathbb{R}^{m+n}} f(\bar{z}) d\bar{z}.$$

**Proposition 4.4** (Rogers Formula; [14, Theorem 5]) *If  $F : \mathbb{R}^{m+n} \times \mathbb{R}^{m+n} \rightarrow \mathbb{R}$  is a non-negative measurable function, then*

$$\begin{aligned} & \int_{\mathcal{X}} \left( \sum_{\bar{z}_1, \bar{z}_2 \in P(\mathbb{Z}^{n+m})} F(g\bar{z}_1, g\bar{z}_2) \right) d\mu_{\mathcal{X}}(g\Gamma) \\ &= \zeta(m+n)^{-2} \int_{\mathbb{R}^{m+n} \times \mathbb{R}^{m+n}} F(\bar{z}_1, \bar{z}_2) d\bar{z}_1 d\bar{z}_2 \\ &+ \zeta(m+n)^{-1} \int_{\mathbb{R}^{m+n}} F(\bar{z}, \bar{z}) d\bar{z} \\ &+ \zeta(m+n)^{-1} \int_{\mathbb{R}^{m+n}} F(\bar{z}, -\bar{z}) d\bar{z}, \end{aligned}$$

where  $P(\mathbb{Z}^{m+n})$  denotes the set of primitive integral vectors in  $\mathbb{Z}^{m+n}$ , and  $\zeta$  denotes Riemann's  $\zeta$ -function.

## 4.2 Non-divergence estimates

We retain the notation from Sect. 2. Given

$$0 < w_1, \dots, w_m < n \quad \text{and} \quad w_1 + \dots + w_m = n,$$

we denote by  $a$  the self-map on  $\mathcal{X}$  induced by

$$a = \text{diag}(e^{w_1}, \dots, e^{w_m}, e^{-1}, \dots, e^{-1}). \quad (4.1)$$

Our goal in this subsection is to analyze the escape of mass for the submanifolds  $a^s \mathcal{Y}$  and bound the Siegel transforms  $\hat{f}(a^s y)$  for  $y \in \mathcal{Y}$ . The following proposition will play a very important role in our arguments.

**Proposition 4.5** *There exists  $\kappa > 0$  such that for every  $L \geq 1$  and  $s \geq \kappa \log L$ ,*

$$\mu_{\mathcal{Y}}(\{y \in \mathcal{Y} : \alpha(a^s y) \geq L\}) \ll_p L^{-p} \quad \text{for all } p < m + n.$$

**Proof** Let  $\chi_L$  be the characteristic function of the subset  $\{\alpha < L\}$  of  $\mathcal{X}$ . By Mahler's Compactness Criterion,  $\chi_L$  has a compact support. We further pick a non-negative  $\rho \in C_c^\infty(G)$  with  $\int_G \rho \, dm_G = 1$ . Let

$$\eta_L(x) := (\rho * \chi_L)(x) = \int_G \rho(g) \chi_L(g^{-1}x) \, dm_G(g), \quad x \in \mathcal{X}.$$

Since  $\mu_{\mathcal{X}}$  is  $G$ -invariant,

$$\int_{\mathcal{X}} \eta_L \, d\mu_{\mathcal{X}} = \int_{\mathcal{X}} \chi_L \, d\mu_{\mathcal{X}} = \mu_{\mathcal{X}}(\{\alpha < L\}).$$

It follows from invariance of  $m_G$  that if  $\mathcal{D}_Z$  is a differential operator as defined as in (2.4), then  $\mathcal{D}_Z \eta_L = (\mathcal{D}_Z \rho) * \chi_L$ . Hence,  $\eta_L \in C_c^\infty(\mathcal{X})$ , and  $\|\eta_L\|_{C^k} \ll \|\rho\|_{C^k}$ .

Note that there exists  $c > 1$  such that for every  $g \in \text{supp}(\rho)$  and all  $x \in \mathcal{X}$ , we have  $\alpha(g^{-1}x) \geq c^{-1}\alpha(x)$ , and thus  $\{\alpha \circ g^{-1} < L\} \subset \{\alpha < cL\}$  and  $\eta_L \leq \chi_{cL}$ . This implies the lower bound

$$\mu_{\mathcal{Y}}(\{y \in \mathcal{Y} : \alpha(a^s y) < cL\}) = \int_{\mathcal{Y}} \chi_{cL}(a^s y) \, d\mu_{\mathcal{Y}}(y) \geq \int_{\mathcal{Y}} \eta_L(a^s y) \, d\mu_{\mathcal{Y}}(y).$$

By Corollary 3.3, there exists  $\delta > 0$  and  $k \geq 1$  such that

$$\begin{aligned} \int_{\mathcal{Y}} \eta_L(a^s y) \, d\mu_{\mathcal{Y}}(y) &= \int_{\mathcal{X}} \eta_L \, d\mu_{\mathcal{X}} + O(e^{-\delta s} \|\eta_L\|_{C^k}) \\ &= \mu_{\mathcal{X}}(\{\alpha < L\}) + O(e^{-\delta s}), \end{aligned}$$



and by Proposition 4.2,

$$\mu_{\mathcal{X}}(\{\alpha \geq L\}) \ll_p L^{-p} \quad \text{for all } p < m + n.$$

Combining these bounds, we get

$$\mu_{\mathcal{Y}}(\{y \in \mathcal{Y} : \alpha(a^s y) < cL\}) \geq \mu_{\mathcal{X}}(\{\alpha < L\}) + O(e^{-\delta s}) = 1 + O_p(L^{-p} + e^{-\delta s}),$$

and thus

$$\mu_{\mathcal{Y}}(\{y \in \mathcal{Y} : \alpha(a^s y) \geq cL\}) \ll_p L^{-p} + e^{-\delta s}.$$

By taking  $s \geq \kappa \log L$  where  $\kappa = \frac{p}{\delta}$ , the proof is finished.  $\square$

**Proposition 4.6** *Let  $f$  be a bounded measurable function on  $\mathbb{R}^{m+n}$  with compact support contained in the open set  $\{(x_{m+1}, \dots, x_{m+n}) \neq 0\}$ . Then, with  $a$  as in (4.1),*

$$\sup_{s \geq 0} \int_{\mathcal{Y}} |\hat{f} \circ a^s| d\mu_{\mathcal{Y}} < \infty.$$

**Remark 4.7** In [4, Theorem 3.2], the authors (implicitly) show a similar uniform estimate for integrals of smooth Siegel transforms over  $\mathrm{SO}(m+n)$ -orbits of a lattice. Their proof is quite different from ours.

**Proof** We note that there exist  $0 < \nu_1 < \nu_2$  and  $\vartheta > 0$  such that the support of  $f$  is contained in the set

$$\{(\bar{x}, \bar{y}) \in \mathbb{R}^{m+n} : \nu_1 \leq \|\bar{y}\| \leq \nu_2, |x_i| \leq \vartheta \|\bar{y}\|^{-w_i}, i = 1, \dots, n\}, \quad (4.2)$$

and without loss of generality we may assume that  $f$  is the characteristic function of this set. We recall that  $\mathcal{Y}$  can be identified with the collection of lattices

$$\{\Lambda_u : u = (u_{ij} : i = 1, \dots, m, j = 1, \dots, n) \in [0, 1)^{m \times n}\}.$$

We set  $\bar{u}_i = (u_{i1}, \dots, u_{in})$ . Then by the definition of the Siegel transform,

$$\hat{f}(a^s \Lambda_u) = \sum_{(\bar{p}, \bar{q}) \in \mathbb{Z}^{m+n} \setminus \{0\}} f(e^{w_1 s}(p_1 + \langle \bar{u}_1, \bar{q} \rangle), \dots, e^{w_m s}(p_m + \langle \bar{u}_m, \bar{q} \rangle), e^{-s} \bar{q}).$$

Denoting by  $\chi_{\bar{q}}^{(i)}$  the characteristic function of the interval  $[-\vartheta \|\bar{q}\|^{-w_i}, \vartheta \|\bar{q}\|^{-w_i}]$ , we rewrite this sum as

$$\begin{aligned} \hat{f}(a^s \Lambda_u) &= \sum_{\nu_1 e^s \leq \|\bar{q}\| \leq \nu_2 e^s} \sum_{\bar{p} \in \mathbb{Z}^m} \prod_{i=1}^m \chi_{\bar{q}}^{(i)}(p_i + \langle \bar{u}_i, \bar{q} \rangle) \\ &= \sum_{\nu_1 e^s \leq \|\bar{q}\| \leq \nu_2 e^s} \prod_{i=1}^m \left( \sum_{p_i \in \mathbb{Z}} \chi_{\bar{q}}^{(i)}(p_i + \langle \bar{u}_i, \bar{q} \rangle) \right). \end{aligned} \quad (4.3)$$

Hence,

$$\int_{\mathcal{Y}} (\hat{f} \circ a^s) d\mu_{\mathcal{Y}} = \sum_{v_1 e^s \leq \|\bar{q}\| \leq v_2 e^s} \prod_{i=1}^m \left( \sum_{p_i \in \mathbb{Z}} \int_{[0,1]^n} \chi_{\bar{q}}^{(i)}(p_i + \langle \bar{u}_i, \bar{q} \rangle) d\bar{u}_i \right).$$

We observe that for each  $i$  and  $p_i \in \mathbb{Z}$ , the volume of the set

$$\{\bar{u} \in [0, 1]^n : |p + \langle \bar{u}, \bar{q} \rangle| \leq \vartheta \|\bar{q}\|^{-w_i}\}$$

is estimated from above by

$$\frac{2\vartheta \|\bar{q}\|^{-w_i}}{\max_j |q_j|} \ll \|\bar{q}\|^{-1-w_i},$$

and we note that the set is empty whenever  $|p| > \sum_j |q_j| + \vartheta \|\bar{q}\|^{-w_i}$ . In particular, it is non-empty for at most  $O(\|\bar{q}\|)$  choices of  $p \in \mathbb{Z}$ . Hence, we deduce that

$$\int_{\mathcal{Y}} (\hat{f} \circ a^s) d\mu_{\mathcal{Y}} \ll \sum_{v_1 e^s \leq \|\bar{q}\| \leq v_2 e^s} \prod_{i=1}^m \|\bar{q}\|^{-w_i} = \sum_{v_1 e^s \leq \|\bar{q}\| \leq v_2 e^s} \|\bar{q}\|^{-n} \ll 1,$$

uniformly in  $s$ . This completes the proof.  $\square$

**Proposition 4.8** *Let  $f$  be a bounded measurable function on  $\mathbb{R}^{m+n}$  with compact support contained in the open set  $\{(x_{m+1}, \dots, x_{m+n}) \neq 0\}$ . Then*

$$\sup_{s \geq 0} (1+s)^{-v_m} \left\| \hat{f} \circ a^s \right\|_{L^2(\mathcal{Y})} < \infty,$$

where  $v_1 = 1$  and  $v_m = 0$  when  $m \geq 2$ .

**Proof** As in the proof of Proposition 4.6, it is sufficient to consider the case when  $f$  is the characteristic function of the set (4.2). Then  $\hat{f}(a^s y)$  is given by (4.3), and we get

$$\begin{aligned} \left\| \hat{f} \circ a^s \right\|_{L^2(\mathcal{Y})}^2 &= \int_{\mathcal{Y}} \hat{f}(a^s y) \hat{f}(a^s y) d\mu_{\mathcal{Y}}(y) \\ &= \sum_{v_1 e^s \leq \|\bar{q}\|, \|\bar{\ell}\| \leq v_2 e^s} \prod_{i=1}^m \left( \sum_{p_i, r_i \in \mathbb{Z}} \int_{[0,1]^n} \chi_{\bar{q}}^{(i)}(p_i + \langle \bar{u}_i, \bar{q} \rangle) \chi_{\bar{\ell}}^{(i)}(r_i + \langle \bar{u}_i, \bar{\ell} \rangle) d\bar{u}_i \right). \end{aligned}$$

For fixed  $\bar{q}, \bar{\ell} \in \mathbb{Z}^n$ , we wish to estimate

$$I_i(\bar{q}, \bar{\ell}) := \sum_{p, r \in \mathbb{Z}} \int_{[0,1]^n} \chi_{\bar{q}}^{(i)}(p + \langle \bar{u}, \bar{q} \rangle) \chi_{\bar{\ell}}^{(i)}(r + \langle \bar{u}, \bar{\ell} \rangle) d\bar{u}.$$

First, we consider the case when  $\bar{q}$  and  $\bar{\ell}$  are linearly independent. Then there exist indices  $j, k = 1, \dots, n$  such that  $q_j \ell_k - q_k \ell_j \neq 0$ . Let us consider the function  $\psi$  on  $\mathbb{R}^2$  defined by  $\psi(x_1, x_2) = \chi_{\bar{q}}^{(i)}(x_1) \chi_{\bar{\ell}}^{(i)}(x_2)$  as well as the periodized function  $\bar{\psi}$  on  $\mathbb{R}^2/\mathbb{Z}^2$  defined by  $\bar{\psi}(x) = \sum_{z \in \mathbb{Z}^2} \psi(z + x)$ . If we set

$$\omega := \sum_{\zeta \neq j, k} q_{\zeta} u_{\zeta} \quad \text{and} \quad \rho := \sum_{\zeta \neq j, k} \ell_{\zeta} u_{\zeta},$$

then we denote by  $S$  the affine map

$$S : (x_1, x_2) \mapsto (q_j x_1 + q_k x_2 + \omega, \ell_j x_1 + \ell_k x_2 + \rho),$$

which induces an affine endomorphism of the torus  $\mathbb{R}^2/\mathbb{Z}^2$ . We note that

$$\sum_{p, r \in \mathbb{Z}} \int_{[0, 1]^2} \chi_{\bar{q}}^{(i)}(p + \langle \bar{u}, \bar{q} \rangle) \chi_{\bar{\ell}}^{(i)}(r + \langle \bar{u}, \bar{\ell} \rangle) du_j du_k = \int_{\mathbb{R}^2/\mathbb{Z}^2} \bar{\psi}(Sx) d\mu(x),$$

where  $\mu$  denotes the Lebesgue probability measure on the torus  $\mathbb{R}^2/\mathbb{Z}^2$ . Since the endomorphism  $S$  preserves  $\mu$ , we see that

$$\int_{\mathbb{R}^2/\mathbb{Z}^2} \bar{\psi}(Sx) d\mu(x) = \int_{\mathbb{R}^2/\mathbb{Z}^2} \bar{\psi}(x) d\mu(x) = \int_{\mathbb{R}^2} \psi(x) dx = 4\vartheta^2 \|\bar{q}\|^{-w_i} \|\bar{\ell}\|^{-w_i}.$$

Therefore, we conclude that in this case,

$$I_i(\bar{q}, \bar{\ell}) \ll \|\bar{q}\|^{-w_i} \|\bar{\ell}\|^{-w_i}. \quad (4.4)$$

Let us now we consider the second case when  $\bar{q}$  and  $\bar{\ell}$  are linearly dependent. Upon re-arranging indices if needed, we may assume that

$$|q_1| = \max(|q_1|, \dots, |q_n|, |\ell_1|, \dots, |\ell_n|). \quad (4.5)$$

In particular,  $q_1 \neq 0$ , and thus  $\ell_1 \neq 0$ , since  $\bar{q}$  and  $\bar{\ell}$  are linearly dependent, so we can define the new variables

$$v_1 = \sum_{j=1}^n (q_j/q_1) u_j = \sum_{j=1}^n (\ell_j/\ell_1) u_j \quad \text{and} \quad v_2 = u_2, \dots, v_n = u_n,$$

and thus

$$I_i(\bar{q}, \bar{\ell}) \leq J_i(\bar{q}, \bar{\ell})$$

where

$$J_i(\bar{q}, \bar{\ell}) := \sum_{p, r \in \mathbb{Z}} \int_{-n}^n \chi_{\bar{q}}^{(i)}(p + q_1 v_1) \chi_{\bar{\ell}}^{(i)}(r + \ell_1 v_1) dv_1.$$

We note that the last integral is non-zero only when  $|p| \ll |q_1|$  and  $|r| \ll |\ell_1|$ . We set  $q_1 = q'd$  and  $\ell_1 = \ell'd$  where  $d = \gcd(q_1, \ell_1)$ . Then  $q_1 r - \ell_1 p = jd$  for some  $j \in \mathbb{Z}$ . We observe that when  $j$  is fixed, then the integers  $p$  and  $r$  satisfy the equation  $q'r - \ell'p = j$ . Since  $\gcd(q', \ell') = 1$ , all the solutions of this equation are of the form  $p = p_0 + kq', r = r_0 + k\ell'$  for  $k \in \mathbb{Z}$ . In particular, it follows that the number of such solutions satisfying  $|p| \ll |q_1|$  and  $|r| \ll |\ell_1|$  is at most  $O(d)$ . We write

$$J_i(\bar{q}, \bar{\ell}) = J_i^{(1)}(\bar{q}, \bar{\ell}) + J_i^{(2)}(\bar{q}, \bar{\ell}),$$

where the first sum is taken over those  $p, r$  with  $q_1 r - \ell_1 p \neq 0$ , and the second sum is taken over those  $p, r$  with  $q_1 r - \ell_1 p = 0$ .

Upon applying a linear change of variables, we obtain

$$\begin{aligned} J_i^{(1)}(\bar{q}, \bar{\ell}) &= \sum_{p, r: q_1 r - \ell_1 p \neq 0} \int_{-n+p/q_1}^{n+p/q_1} \chi_{\bar{q}}^{(i)}(q_1 v_1) \chi_{\bar{\ell}}^{(i)}((q_1 r - \ell_1 p)/q_1 + \ell_1 v_1) dv_1 \\ &\ll d \sum_{j \in \mathbb{Z} \setminus \{0\}} \int_{-\infty}^{\infty} \chi_{\bar{q}}^{(i)}(q_1 v_1) \chi_{\bar{\ell}}^{(i)}(jd/q_1 + \ell_1 v_1) dv_1. \end{aligned}$$

Let us consider the function

$$\rho_i(x) := \int_{-\infty}^{\infty} \chi_{\bar{q}}^{(i)}(q_1 v_1) \chi_{\bar{\ell}}^{(i)}(xd/q_1 + \ell_1 v_1) dv_1.$$

We note that the integrand equals the indicator function of the intersection of the intervals

$$\left[ -\vartheta |q_1|^{-1} \|\bar{q}\|^{-w_i}, \vartheta |q_1|^{-1} \|\bar{q}\|^{-w_i} \right]$$

and

$$\left[ -xd/(q_1 \ell_1) - \vartheta |\ell_1|^{-1} \|\bar{\ell}\|^{-w_i}, -xd/(q_1 \ell_1) + \vartheta |\ell_1|^{-1} \|\bar{\ell}\|^{-w_i} \right],$$

and thus it follows that  $\rho_i$  is non-increasing when  $x \geq 0$ , and non-decreasing when  $x \leq 0$ . This implies that

$$\sum_{j \in \mathbb{Z} \setminus \{0\}} \rho_i(j) \leq \int_{-\infty}^{\infty} \rho_i(x) dx.$$

Since

$$\int_{-\infty}^{\infty} \rho_i(x) dx = \left( \int_{-\infty}^{\infty} \chi_{\bar{q}}^{(i)}(q_1 v_1) dv_1 \right) \left( \int_{-\infty}^{\infty} \chi_{\bar{\ell}}^{(i)}(xd/q_1) dx \right) \ll d^{-1} \|\bar{q}\|^{-w_i} \|\bar{\ell}\|^{-w_i},$$

we conclude that

$$J_i^{(1)}(\bar{q}, \bar{\ell}) \ll d \sum_{j \in \mathbb{Z} \setminus \{0\}} \rho_i(j) \ll \|\bar{q}\|^{-w_i} \|\bar{\ell}\|^{-w_i}.$$

Next, we proceed with estimation of  $J_i^{(2)}(\bar{q}, \bar{\ell})$ . Let  $c_0 := \min\{\|\bar{q}\| : \bar{q} \in \mathbb{Z}^n \setminus \{0\}\}$ . We set

$$N(q_1, \ell_1) = \left| \{(p, r) \in \mathbb{Z}^2 : q_1 r = \ell_1 p \text{ with } |p| \leq (c_0^{-n} \vartheta + n) |q_1|\} \right|$$

and obtain

$$\begin{aligned} J_i^{(2)}(\bar{q}, \bar{\ell}) &= \sum_{p, r: q_1 r - \ell_1 p = 0} \int_{-n+p/q_1}^{n+p/q_1} \chi_{\bar{q}}^{(i)}(q_1 v_1) \chi_{\bar{\ell}}^{(i)}(\ell_1 v_1) dv_1 \\ &\leq N(q_1, \ell_1) \int_{-\infty}^{\infty} \chi_{\bar{q}}^{(i)}(q_1 v_1) \chi_{\bar{\ell}}^{(i)}(\ell_1 v_1) dv_1 \\ &\ll N(q_1, \ell_1) |q_1|^{-1} \|\bar{q}\|^{-w_i} \ll N(q_1, \ell_1) \max(\|\bar{q}\|, \|\bar{\ell}\|)^{-(1+w_i)}, \end{aligned}$$

where we used that  $q_1$  is chosen according to (4.5). Combining the obtained estimates for  $J_i^{(1)}(\bar{q}, \bar{\ell})$  and  $J_i^{(2)}(\bar{q}, \bar{\ell})$ , we conclude that when  $\bar{q}$  and  $\bar{\ell}$  are linearly dependent,

$$J_i(\bar{q}, \bar{\ell}) \ll \|\bar{q}\|^{-w_i} \|\bar{\ell}\|^{-w_i} + N(q_1, \ell_1) \max(\|\bar{q}\|, \|\bar{\ell}\|)^{-(1+w_i)}, \quad (4.6)$$

where  $q_1$  is chosen according to (4.5).

Now we proceed to estimate

$$\left\| \hat{f} \circ a^s \right\|_{L^2(\mathcal{Y})}^2 \ll \sum_{v_1 e^s \leq \|\bar{q}\|, \|\bar{\ell}\| \leq v_2 e^s} \prod_{i=1}^m I_i(\bar{q}, \bar{\ell}). \quad (4.7)$$

Using (4.4), the sum in (4.7) over linearly independent  $\bar{q}$  and  $\bar{\ell}$  can be estimated as

$$\ll \sum_{v_1 e^s \leq \|\bar{q}\|, \|\bar{\ell}\| \leq v_2 e^s} \prod_{i=1}^m \|\bar{q}\|^{-w_i} \|\bar{\ell}\|^{-w_i} \ll \sum_{v_1 e^s \leq \|\bar{q}\|, \|\bar{\ell}\| \leq v_2 e^s} \|\bar{q}\|^{-n} \|\bar{\ell}\|^{-n} \ll 1.$$

For a subset  $I$  of  $\{1, \dots, m\}$ , we set  $w(I) := \sum_{i \in I} w_i$ . Then using (4.6), we deduce that the sum in (4.7) over linearly dependent  $\bar{q}$  and  $\bar{\ell}$  is bounded by

$$\begin{aligned} &\ll \sum_{v_1 e^s \leq \|\bar{q}\|, \|\bar{\ell}\| \leq v_2 e^s}^* \sum_{I \subset \{1, \dots, m\}} \|\bar{q}\|^{-w(I)} \|\bar{\ell}\|^{-w(I)} N(q_1, \ell_1)^{|I^c|} \\ &\quad \times \max(\|\bar{q}\|, \|\bar{\ell}\|)^{-(|I^c| + w(I^c))} \end{aligned}$$

$$\ll \sum_{I \subset \{1, \dots, m\}} (e^s)^{-(n+|I^c|+w(I))} \sum_{v_1 e^s \leq \|\bar{q}\|, \|\bar{\ell}\| \leq v_2 e^s}^* N(q_1, \ell_1)^{|I^c|}. \quad (4.8)$$

The star indicates that the sum is taken over linearly dependent  $\bar{q}$  and  $\bar{\ell}$ .

When  $I^c = \emptyset$ , then  $w(I) = n$ . Since the number of  $(\bar{q}, \bar{\ell})$  satisfying  $v_1 e^s \leq \|\bar{q}\|, \|\bar{\ell}\| \leq v_2 e^s$  is estimated as  $O(e^{2ns})$ , it is clear that the corresponding term in the above sum is uniformly bounded.

Now we suppose that  $I^c \neq \emptyset$ . Since  $\bar{q}$  and  $\bar{\ell}$  are linearly dependent, the vector  $\bar{\ell}$  is uniquely determined given  $\ell_1$  and  $\bar{q}$ , and we obtain that for some  $v'_1, v'_2 > 0$ ,

$$\sum_{v_1 e^s \leq \|\bar{q}\|, \|\bar{\ell}\| \leq v_2 e^s}^* N(q_1, \ell_1)^{|I^c|} \ll (e^s)^{n-1} \sum_{v'_1 e^s \leq |q_1| \leq v'_2 e^s} \sum_{1 \leq |\ell_1| \leq |q_1|} N(q_1, \ell_1)^{|I^c|}.$$

We shall use the following lemma:

**Lemma 4.9** *For every  $k \geq 1$ ,*

$$\sum_{1 \leq q \leq T} \sum_{1 \leq \ell \leq q} N(q, \ell)^k \ll T^{k+1} (\log T)^{v_k}$$

where  $v_1 = 1$  and  $v_k = 0$  when  $k \geq 2$ .

**Proof** We observe that the sum of  $N(q, \ell)^k$  over  $1 \leq \ell \leq q$  is equal to the number of solutions  $(p_1, \dots, p_k, r_1, \dots, r_k, \ell)$  of the system of equations

$$qr_1 - \ell p_1 = 0, \dots, qr_k - \ell p_k = 0 \quad (4.9)$$

satisfying

$$|p_1|, \dots, |p_k| \leq (c_0^{-n} \vartheta + n)q \quad \text{and} \quad 1 \leq \ell \leq q.$$

We order these solutions according to  $d := \gcd(q, \ell)$ . Let  $q = q'd$  and  $\ell = \ell'd$ . Then  $q'$  and  $\ell'$  are coprime, and the system (4.9) is equivalent to

$$q'r_1 - \ell'p_1 = 0, \dots, q'r_k - \ell'p_k = 0. \quad (4.10)$$

Because of coprimality, each  $p_i$  have to be divisible by  $q'$ , so that the number of such  $p_i$ 's is at most  $O(q/q') = O(d)$ . We note that given  $d$  the number of possible choices for  $\ell$  is at most  $q/d$ , and  $(p_1, \dots, p_k, \ell)$  uniquely determine  $(r_1, \dots, r_k)$ . Hence, the number of solutions of (4.9) is estimated by

$$\ll \sum_{d|q} (q/d) d^k = q \sigma_{k-1}(q),$$

where  $\sigma_{k-1}(q) = \sum_{d|q} d^{k-1}$ . Writing  $q = q'd$ , we conclude that

$$\begin{aligned} \sum_{1 \leq q \leq T} \sum_{1 \leq \ell \leq q} N(q, \ell)^k &\ll \sum_{1 \leq q \leq T} q \sigma_{k-1}(q) \leq T \sum_{1 \leq q' \leq T} \sum_{q=1}^{\lfloor T/q' \rfloor} d^{k-1} \\ &\ll T^{k+1} \sum_{1 \leq q' \leq T} (q')^{-k} \ll T^{k+1} (\log T)^{v_k}. \end{aligned}$$

This proves the lemma.  $\square$

**Remark 4.10** A similar estimate in the case  $k = 1$  was proved by Schmidt in [16].

A simple modification of this argument also gives that

$$\sum_{1 \leq |q| \leq T} \sum_{1 \leq |\ell| \leq |q|} N(q, \ell)^k \ll T^{k+1} (\log T)^{v_k}.$$

Hence, it follows that

$$\sum_{\substack{* \\ v_1 e^s \leq \|\bar{q}\|, \|\bar{\ell}\| \leq v_2 e^s}} N(q_1, \ell_1)^{|I^c|} \ll (e^s)^{n+|I^c|} (1+s)^{v(I)},$$

where  $v(I) = 1$  when  $|I^c| = 1$  and  $v(I) = 0$  otherwise. Thus, the sum (4.8) is estimated as

$$\ll 1 + \sum_{I \subsetneq \{1, \dots, m\}} e^{-sw(I)} (1+s)^{v(I)}.$$

The terms in this sum are uniformly bounded unless  $I = \emptyset$  and  $|I^c| = 1$ , namely, when  $m > 1$ . When  $m = 1$ , we obtain the bound  $O(1+s)$ . This proves the proposition.  $\square$

### 4.3 Truncated Siegel transform

The Siegel transform of a compactly supported function is typically unbounded on  $\mathcal{X}$ ; to deal with this complication, it is natural to approximate  $\hat{f}$  by compactly supported functions on  $\mathcal{X}$ , the so called *truncated Siegel transforms*, which we shall denote by  $\hat{f}^{(L)}$ . They will be constructed using a smooth cut-off function  $\eta_L$ , which will be defined in the following lemma.

**Lemma 4.11** *For every  $c > 1$ , there exists a family  $(\eta_L)$  in  $C_c^\infty(\mathcal{X})$  satisfying:*

$$0 \leq \eta_L \leq 1, \quad \eta_L = 1 \text{ on } \{\alpha \leq c^{-1} L\}, \quad \eta_L = 0 \text{ on } \{\alpha > c L\}, \quad \|\eta_L\|_{C^k} \ll 1.$$

**Proof** Let  $\chi_L$  denote the indicator function of the subset  $\{\alpha \leq L\} \subset \mathcal{X}$ , and pick a non-negative  $\phi \in C_c^\infty(G)$  with  $\int_G \phi \, dm_G = 1$  and with support in a sufficiently small neighbourhood of identity in  $G$  to ensure that for all  $g \in \text{supp}(\phi)$  and  $x \in \mathcal{X}$ ,

$$c^{-1} \alpha(x) \leq \alpha(g^{\pm 1} x) \leq c \alpha(x).$$

We now define  $\eta_L$  as

$$\eta_L(x) := (\phi * \chi_L)(x) = \int_G \phi(g) \chi_L(g^{-1} x) \, dm_G(g).$$

Since  $\phi \geq 0$  and  $\int_G \phi \, dm_G = 1$ , it is clear that  $0 \leq \eta_L \leq 1$ . If  $\alpha(x) \leq c^{-1} L$ , then for  $g \in \text{supp}(\phi)$ , we have  $\alpha(g^{-1} x) \leq L$ , so that  $\eta_L(x) = \int_G \phi \, dm_G = 1$ . If  $\alpha(x) > c L$ , then for  $g \in \text{supp}(\phi)$ , we have  $\alpha(g^{-1} x) > L$ , so that  $\eta_L(x) = 0$ .

To prove the last property, we observe that it follows from invariance of  $m_G$  that for a differential operator  $\mathcal{D}_Z$  as in (2.4), we have  $\mathcal{D}_Z \eta_L = (\mathcal{D}_Z \phi) * \chi_L$ . Therefore,  $\text{supp}(\mathcal{D}_Z \eta_L) \subset \{\alpha \leq c L\}$  and  $\|\mathcal{D}_Z \eta_L\|_{C^0} \leq \|\mathcal{D}_Z \phi\|_{L^1(G)}$ , whence  $\|\eta_L\|_{C^k} \ll 1$ .  $\square$

For a bounded function  $f : \mathbb{R}^{m+n} \rightarrow \mathbb{R}$  with compact support, we define the *truncated Siegel transform* of  $f$  as

$$\hat{f}^{(L)} := \hat{f} \cdot \eta_L.$$

We record some basic properties of this transform that will be used later in the proofs.

**Lemma 4.12** *For  $f \in C_c^\infty(\mathbb{R}^{m+n})$ , the truncated Siegel transform  $\hat{f}^{(L)}$  is in  $C_c^\infty(\mathcal{X})$ , and it satisfies*

$$\|\hat{f}^{(L)}\|_{L^p(\mathcal{X})} \leq \|\hat{f}\|_{L^p(\mathcal{X})} \ll_{\text{supp}(f), p} \|f\|_{C^0} \quad \text{for all } p < m+n, \quad (4.11)$$

$$\|\hat{f}^{(L)}\|_{C^0} \ll_{\text{supp}(f)} L \|f\|_{C^0}, \quad (4.12)$$

$$\|\hat{f}^{(L)}\|_{C^k} \ll_{\text{supp}(f)} L \|f\|_{C^k}, \quad (4.13)$$

$$\|\hat{f} - \hat{f}^{(L)}\|_{L^1(\mathcal{X})} \ll_{\text{supp}(f), \tau} L^{-\tau} \|f\|_{C^0} \quad \text{for all } \tau < m+n-1, \quad (4.14)$$

$$\|\hat{f} - \hat{f}^{(L)}\|_{L^2(\mathcal{X})} \ll_{\text{supp}(f), \tau} L^{-(\tau-1)/2} \|f\|_{C^0} \quad \text{for all } \tau < m+n-1. \quad (4.15)$$

Moreover, the implied constants are uniform when  $\text{supp}(f)$  is contained in a fixed compact set.

**Proof** It follows from Proposition 4.1 that

$$|\hat{f}^{(L)}| \ll_{\text{supp}(f)} \|f\|_{C^0} \alpha \eta_L.$$

Since  $0 \leq \eta_L \leq 1$ , (4.11) follows from Proposition 4.2, and the upper bound in (4.12) holds since  $\text{supp}(\eta_L) \subset \{\alpha \leq cL\}$ .



We observe that for a differential operator  $\mathcal{D}_Z$  as in (2.4), we have  $\mathcal{D}_Z(\hat{f}) = \widehat{\mathcal{D}_Z f}$ . Hence, we deduce from Proposition 4.1 that

$$\left| \mathcal{D}_Z(\hat{f}) \right| \ll_{\text{supp}(f)} \|f\|_{C^k} \alpha.$$

Since  $\text{supp}(\eta_L) \subset \{\alpha \leq cL\}$  and  $\|\eta_L\|_{C^k} \ll 1$ , we deduce that

$$\left\| \hat{f}^{(L)} \right\|_{C^k} \ll_{\text{supp}(f)} L \|f\|_{C^k}$$

proving (4.13).

To prove (4.14), we observe that since  $0 \leq \eta_L \leq 1$  and  $\eta_L = 1$  on  $\{\alpha < c^{-1}L\}$ , it follows from Proposition 4.1 that

$$\left\| \hat{f} - \hat{f}^{(L)} \right\|_{L^1(\mathcal{X})} = \int_{\mathcal{X}} |\hat{f}| \cdot |1 - \eta_L| d\mu_{\mathcal{X}} \ll_{\text{supp}(f)} \|f\|_{C^0} \int_{\{\alpha \geq c^{-1}L\}} \alpha d\mu_{\mathcal{X}}.$$

Hence, applying the Hölder inequality with  $1 \leq p < m+n$  and  $q = (1 - 1/p)^{-1}$ , we deduce that

$$\left\| \hat{f} - \hat{f}^{(L)} \right\|_{L^1(\mathcal{X})} \ll_{\text{supp}(f)} \|f\|_{C^0} \|\alpha\|_p \mu_{\mathcal{X}}(\{\alpha \geq c^{-1}L\})^{1/q}.$$

Now it follows from Proposition 4.2 that

$$\left\| \hat{f} - \hat{f}^{(L)} \right\|_{L^1(\mathcal{X})} \ll_{\text{supp}(f), p} \|f\|_{C^0} L^{-(p-1)},$$

which proves (4.14). The proof of (4.15) is similar, and we omit the details.  $\square$

## 5 CLT for smooth Siegel transforms

Assume that  $f \in C_c^\infty(\mathbb{R}^{m+n})$  satisfies  $f \geq 0$  and  $\text{supp}(f) \subset \{(x_{m+1}, \dots, x_{m+n}) \neq 0\}$ . We shall in this section analyze the asymptotic behavior of the averages

$$F_N(y) := \frac{1}{\sqrt{N}} \sum_{s=0}^{N-1} \left( \hat{f}(a^s y) - \mu_{\mathcal{Y}}(\hat{f} \circ a^s) \right) \quad \text{with } y \in \mathcal{Y},$$

and prove the following result:

**Theorem 5.1** *If  $m \geq 2$  (and thus  $m+n \geq 3$ ) and  $f$  is as above, then the variance*

$$\begin{aligned} \sigma_f^2 &:= \sum_{s=-\infty}^{\infty} \left( \int_{\mathcal{X}} (\hat{f} \circ a^s) \hat{f} d\mu_{\mathcal{X}} - \mu_{\mathcal{X}}(\hat{f})^2 \right) \\ &= \zeta(m+n)^{-1} \sum_{s=-\infty}^{\infty} \sum_{p, q \geq 1} \left( \int_{\mathbb{R}^{m+n}} f(pa^s z) f(qz) dz + \int_{\mathbb{R}^{m+n}} f(pa^s z) f(-qz) dz \right). \end{aligned}$$

is finite, and for every  $\xi \in \mathbb{R}$ ,

$$\mu_{\mathcal{Y}}(\{y \in \mathcal{Y} : F_N(y) < \xi\}) \rightarrow \text{Norm}_{\sigma_f}(\xi) \quad (5.1)$$

as  $N \rightarrow \infty$ .

The proof of Theorem 5.1 follows the same plan as the proof of Theorem 3.1, but we need to develop an additional approximation argument which involves truncations of the Siegel transform  $\hat{f}$ . This can potentially change the behaviour of the averages  $F_N$ , so we will have to take into account possible escapes of masses for the sequences of submanifolds  $a^s \mathcal{Y}$  in  $\mathcal{X}$ .

Throughout the proof, we shall frequently make use of the basic observation that if we approximate  $F_N$  by  $\tilde{F}_N$  in such a way that  $\|F_N - \tilde{F}_N\|_{L^1(\mathcal{Y})} \rightarrow 0$ , then  $F_N$  and  $\tilde{F}_N$  will have the same convergence in distribution. Each time we apply this observation, the new approximation will depend on some sequence which depend on  $N$ ; ultimately, we will end up with three different, but interrelated, sequences  $K(N)$ ,  $L(N)$  and  $M(N)$ , which need to be matched. In Sect. 5.3, we will provide explicit choices for these sequences.

Let

$$\tilde{F}_N := \frac{1}{\sqrt{N}} \sum_{s=M}^{N-1} \left( \hat{f} \circ a^s - \mu_{\mathcal{Y}}(\hat{f} \circ a^s) \right) \quad (5.2)$$

for some  $M = M(N) \rightarrow \infty$  that will be chosen later. We observe that

$$\begin{aligned} \|F_N - \tilde{F}_N\|_{L^1(\mathcal{Y})} &\leq \frac{1}{\sqrt{N}} \sum_{s=0}^{M-1} \left\| \hat{f} \circ a^s - \mu_{\mathcal{Y}}(\hat{f} \circ a^s) \right\|_{L^1(\mathcal{Y})} \\ &\leq \frac{2M}{\sqrt{N}} \sup_{s \geq 0} \int_{\mathcal{Y}} |\hat{f} \circ a^s| d\mu_{\mathcal{Y}}. \end{aligned} \quad (5.3)$$

and thus, by Proposition 4.6 we see that

$$\|F_N - \tilde{F}_N\|_{L^1(\mathcal{Y})} \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

provided

$$M = o(N^{1/2}). \quad (5.4)$$

In particular, it follows that if (5.1) holds for  $\tilde{F}_N$ , then it also holds for  $F_N$ , we shall prove the former. In order to simplify notation, let us drop the tilde, and assume from now on that  $F_N$  is given by (5.2).

Given a sequence  $L = L(N)$ , which shall be chosen later, we consider the average

$$F_N^{(L)} := \frac{1}{\sqrt{N}} \sum_{s=M}^{N-1} \left( \hat{f}^{(L)} \circ a^s - \mu_{\mathcal{Y}}(\hat{f}^{(L)} \circ a^s) \right)$$

defined for the truncated Siegel transforms  $\hat{f}^{(L)}$  introduced in Sect. 4.3. We have

$$\begin{aligned} \|F_N - F_N^{(L)}\|_{L^1(\mathcal{Y})} &\leq \frac{1}{\sqrt{N}} \sum_{s=M}^{N-1} \left\| \left( \hat{f} \circ a^s - \hat{f}^{(L)} \circ a^s \right) - \mu_{\mathcal{Y}} \left( \hat{f} \circ a^s - \hat{f}^{(L)} \circ a^s \right) \right\|_{L^1(\mathcal{Y})} \\ &\leq \frac{2}{\sqrt{N}} \sum_{s=M}^{N-1} \left\| \hat{f} \circ a^s - \hat{f}^{(L)} \circ a^s \right\|_{L^1(\mathcal{Y})}. \end{aligned}$$

We recall that  $\hat{f}^{(L)} = \hat{f} \cdot \eta_L$ ,  $0 \leq \eta_L \leq 1$ , and  $\eta_L(x) = 1$  when  $\alpha(x) \leq c^{-1}L$ , so that

$$\begin{aligned} \left\| \hat{f} \circ a^s - \hat{f}^{(L)} \circ a^s \right\|_{L^1(\mathcal{Y})} &= \left\| (\hat{f} \circ a^s)(1 - \eta_L \circ a^s) \right\|_{L^1(\mathcal{Y})} \\ &\leq \int_{\{\alpha(a^s y) \geq c^{-1}L\}} |\hat{f}(a^s y)| d\mu_{\mathcal{Y}}(y). \end{aligned}$$

Hence, by the Cauchy–Schwarz inequality,

$$\left\| \hat{f} \circ a^s - \hat{f}^{(L)} \circ a^s \right\|_{L^1(\mathcal{Y})} \leq \left\| \hat{f} \circ a^s \right\|_{L^2(\mathcal{Y})} \mu_{\mathcal{Y}} \left( \left\{ y \in \mathcal{Y} : \alpha(a^s y) \geq c^{-1}L \right\} \right)^{1/2}.$$

Let us now additionally assume that

$$M \gg \log L, \quad (5.5)$$

so that the assumption of Proposition 4.5 is satisfied when  $s \geq M$ . This implies that

$$\mu_{\mathcal{Y}} \left( \left\{ y \in \mathcal{Y} : \alpha(a^s y) \geq c^{-1}L \right\} \right) \ll_p L^{-p} \quad \text{for all } p < m+n.$$

We also recall that by Proposition 4.8 when  $m \geq 2$ ,

$$\sup_{s \geq 0} \left\| \hat{f} \circ a^s \right\|_{L^2(\mathcal{Y})} < \infty,$$

whence, for  $s \geq M$ ,

$$\left\| \hat{f} \circ a^s - \hat{f}^{(L)} \circ a^s \right\|_{L^1(\mathcal{Y})} \ll_p L^{-p/2},$$

and thus

$$\left\| F_N - F_N^{(L)} \right\|_{L^1(\mathcal{Y})} \ll_p N^{1/2} L^{-p/2} \quad \text{for all } p < m+n. \quad (5.6)$$

If we now choose  $L = L(N) \rightarrow \infty$  so that

$$N = o(L^p) \quad \text{for some } p < m+n, \quad (5.7)$$

then it follows that

$$\left\| F_N - F_N^{(L)} \right\|_{L^1(\mathcal{Y})} \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

Hence, if we can show Theorem 5.1 for the averages  $F_N^{(L)}$  with the parameter constraints above, then it would also hold for  $F_N$ . In order to prove CLT for  $(F_N^{(L)})$ , we follow the route of Proposition 3.4 and estimate cumulants and  $L^2$ -norms of the sequence.

## 5.1 Estimating cumulants

We set

$$\psi_s^{(L)}(y) := \hat{f}^{(L)}(a^s y) - \mu_{\mathcal{Y}}(\hat{f}^{(L)} \circ a^s).$$

Our aim is to estimate

$$\text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(F_N^{(L)}) = \frac{1}{N^{r/2}} \sum_{s_1, \dots, s_r = M}^{N-1} \text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(\psi_{s_1}^{(L)}, \dots, \psi_{s_r}^{(L)}) \quad (5.8)$$

when  $r \geq 3$ . The argument proceeds as in Sect. 3.2, but we have to refine the previous estimates to take into account the dependence on the parameters  $L$  and  $M$ . Using the notation from Sect. 3.2, we have the decomposition

$$\{M, \dots, N-1\}^r = \Omega(\beta_{r+1}; M, N) \cup \left( \bigcup_{j=0}^r \bigcup_{|\mathcal{Q}| \geq 2} \Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; M, N) \right), \quad (5.9)$$

where

$$\begin{aligned} \Omega(\beta_{r+1}; M, N) &:= \{M, \dots, N-1\}^r \cap \Delta(\beta_{r+1}), \\ \Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; M, N) &:= \{M, \dots, N-1\}^r \cap \Delta_{\mathcal{Q}}(\alpha_j, \beta_{j+1}). \end{aligned}$$

We decompose the sum into the sums over  $\Omega(\beta_{r+1}; M, N)$  and  $\Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; M, N)$ . Let us choose  $M$  so that

$$M > \beta_{r+1}. \quad (5.10)$$

Then  $\Omega(\beta_{r+1}; M, N) = \emptyset$ , and does not contribute to our estimates.

### 5.1.1 Case 1: Summing over $(s_1, \dots, s_r) \in \Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; M, N)$ with $\mathcal{Q} = \{\{0\}, \{1, \dots, r\}\}$

In this case, we shall show that

$$\text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(\psi_{s_1}^{(L)}, \dots, \psi_{s_r}^{(L)}) \approx \text{Cum}_{\mu_{\mathcal{X}}}^{(r)}(\phi^{(L)} \circ a^{s_1}, \dots, \phi^{(L)} \circ a^{s_r}) \quad (5.11)$$

where  $\phi^{(L)} := \hat{f}^{(L)} - \mu_{\mathcal{X}}(\hat{f}^{(L)})$ . This reduces to estimating the integrals

$$\begin{aligned} & \int_{\mathcal{Y}} \left( \prod_{i \in I} \psi_{s_i}^{(L)} \right) d\mu_{\mathcal{Y}} \\ &= \sum_{J \subset I} (-1)^{|I \setminus J|} \left( \int_{\mathcal{Y}} \left( \prod_{i \in J} \hat{f}^{(L)} \circ a^{s_i} \right) d\mu_{\mathcal{Y}} \right) \prod_{i \in I \setminus J} \left( \int_{\mathcal{Y}} (\hat{f}^{(L)} \circ a^{s_i}) d\mu_{\mathcal{Y}} \right). \end{aligned} \quad (5.12)$$

If  $(s_1, \dots, s_r) \in \Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; N)$ , and thus

$$|s_{i_1} - s_{i_2}| \leq \alpha_j \quad \text{and} \quad s_{i_1} \geq \beta_{j+1} \quad \text{for all } 1 \leq i_1, i_2 \leq r,$$

it follows from Corollary 3.3 with  $r = 1$  that there exists  $\delta > 0$  such that

$$\int_{\mathcal{Y}} (\hat{f}^{(L)} \circ a^{s_i}) d\mu_{\mathcal{Y}} = \mu_{\mathcal{X}}(\hat{f}^{(L)}) + O\left(e^{-\delta\beta_{j+1}} \|\hat{f}^{(L)}\|_{C^k}\right). \quad (5.13)$$

For a fixed  $J \subset I$ , we define

$$\Phi^{(L)} := \prod_{i \in J} \hat{f}^{(L)} \circ a^{s_i - s_1},$$

and note that for some  $\xi = \xi(m, n, k) > 0$ , we have

$$\|\Phi^{(L)}\|_{C^k} \ll \prod_{i \in J} \|\hat{f}^{(L)} \circ a^{s_i - s_1}\|_{C^k} \ll e^{|J|\xi\alpha_j} \|\hat{f}^{(L)}\|_{C^k}^{|J|}.$$

If we again apply Corollary 3.3 to the function  $\Phi^{(L)}$ , we obtain

$$\begin{aligned} & \int_{\mathcal{Y}} \left( \prod_{i \in J} \hat{f}^{(L)} \circ a^{s_i} \right) d\mu_{\mathcal{Y}} = \int_{\mathcal{Y}} (\Phi^{(L)} \circ a^{s_1}) d\mu_{\mathcal{Y}} \\ &= \int_{\mathcal{X}} \Phi^{(L)} d\mu_{\mathcal{X}} + O\left(e^{-\delta\beta_{j+1}} \|\Phi^{(L)}\|_{C^k}\right) \\ &= \int_{\mathcal{X}} \left( \prod_{i \in J} \hat{f}^{(L)} \circ a^{s_i} \right) d\mu_{\mathcal{X}} + O\left(e^{-\delta\beta_{j+1}} e^{r\xi\alpha_j} \|\hat{f}^{(L)}\|_{C^k}^{|J|}\right), \end{aligned} \quad (5.14)$$

where we used that  $\mu_{\mathcal{X}}$  is invariant under the transformation  $a$ . Let us now choose the exponents  $\alpha_j$  and  $\beta_{j+1}$  so that  $\delta\beta_{j+1} - r\xi\alpha_j > 0$ . Combining (5.12), (5.13) and (5.14), we deduce that

$$\begin{aligned}
 \int_{\mathcal{Y}} \left( \prod_{i \in I} \psi_{s_i}^{(L)} \right) d\mu_{\mathcal{Y}} &= \sum_{J \subset I} (-1)^{|I \setminus J|} \left( \int_{\mathcal{X}} \left( \prod_{i \in J} \hat{f}^{(L)} \circ a^{s_i} \right) d\mu_{\mathcal{X}} \right) \mu_{\mathcal{X}} \left( \hat{f}^{(L)} \right)^{|I \setminus J|} \\
 &\quad + O \left( e^{-\delta\beta_{j+1}} e^{r\xi\alpha_j} \left\| \hat{f}^{(L)} \right\|_{C^k}^{|I|} \right) \\
 &= \int_{\mathcal{X}} \prod_{i \in I} \left( \hat{f}^{(L)} \circ a^{s_i} - \mu_{\mathcal{X}}(\hat{f}^{(L)}) \right) d\mu_{\mathcal{X}} \\
 &\quad + O \left( e^{-(\delta\beta_{j+1} - r\xi\alpha_j)} \left\| \hat{f}^{(L)} \right\|_{C^k}^{|I|} \right), \tag{5.15}
 \end{aligned}$$

and thus, for any partition  $\mathcal{P}$ ,

$$\begin{aligned}
 \prod_{I \in \mathcal{P}} \int_{\mathcal{Y}} \left( \prod_{i \in I} \psi_{s_i}^{(L)} \right) d\mu_{\mathcal{Y}} &= \prod_{I \in \mathcal{P}} \int_{\mathcal{X}} \left( \prod_{i \in I} \phi^{(L)} \circ a^{s_i} \right) d\mu_{\mathcal{X}} \\
 &\quad + O \left( e^{-(\delta\beta_{j+1} - r\xi\alpha_j)} \left\| \hat{f}^{(L)} \right\|_{C^k}^r \right),
 \end{aligned}$$

and consequently,

$$\begin{aligned}
 \text{Cum}_{\mu_{\mathcal{Y}}}^{(r)} \left( \psi_{s_1}^{(L)}, \dots, \psi_{s_r}^{(L)} \right) &= \text{Cum}_{\mu_{\mathcal{X}}}^{(r)} \left( \phi^{(L)} \circ a^{s_1}, \dots, \phi^{(L)} \circ a^{s_r} \right) \\
 &\quad + O \left( e^{-(\delta\beta_{j+1} - r\xi\alpha_j)} \left\| \hat{f}^{(L)} \right\|_{C^k}^r \right) \tag{5.16}
 \end{aligned}$$

whenever  $(s_1, \dots, s_r) \in \Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; M, N)$  with  $\mathcal{Q} = \{\{0\}, \{1, \dots, r\}\}$ , from which (5.11) follows.

We now claim that

$$\begin{aligned}
 &\left| \text{Cum}_{\mu_{\mathcal{X}}}^{(r)} \left( \phi^{(L)} \circ a^{s_1}, \dots, \phi^{(L)} \circ a^{s_r} \right) \right| \\
 &\ll_f \left\| \hat{f}^{(L)} \right\|_{C^0}^{(r-(m+n-1))^+} \left\| \hat{f}^{(L)} \right\|_{L^{m+n-1}(\mathcal{X})}^{\min(r, m+n-1)}, \tag{5.17}
 \end{aligned}$$

where we use the notation  $x^+ = \max(x, 0)$ . The implied constant in (5.17) and below in the proof depend only on  $\text{supp}(f)$ . By the definition of the cumulant, to prove (5.17), it suffices to show that for every  $z \geq 1$  and indices  $i_1, \dots, i_z$ ,

$$\begin{aligned}
 &\int_{\mathcal{X}} \left| \left( \phi^{(L)} \circ a^{s_{i_1}} \right) \dots \left( \phi^{(L)} \circ a^{s_{i_z}} \right) \right| d\mu_{\mathcal{X}} \\
 &\ll_f \left\| \hat{f}^{(L)} \right\|_{C^0}^{(z-(m+n-1))^+} \left\| \hat{f}^{(L)} \right\|_{L^{m+n-1}(\mathcal{X})}^{\min(z, m+n-1)}. \tag{5.18}
 \end{aligned}$$

Using the generalized Hölder inequality, we deduce that when  $z \leq m+n-1$ ,

$$\int_{\mathcal{X}} \left| \left( \phi^{(L)} \circ a^{s_{i_1}} \right) \dots \left( \phi^{(L)} \circ a^{s_{i_z}} \right) \right| d\mu_{\mathcal{X}}$$

$$\begin{aligned} &\leq \left\| \phi^{(L)} \circ a^{s_{i_1}} \right\|_{L^{m+n-1}(\mathcal{X})} \cdots \left\| \phi^{(L)} \circ a^{s_{i_z}} \right\|_{L^{m+n-1}(\mathcal{X})} \\ &\ll \left\| \hat{f}^{(L)} \right\|_{L^{m+n-1}(\mathcal{X})}^z. \end{aligned}$$

Also when  $z > m + n - 1$ ,

$$\begin{aligned} &\int_{\mathcal{X}} \left| \left( \phi^{(L)} \circ a^{s_{i_1}} \right) \cdots \left( \phi^{(L)} \circ a^{s_{i_z}} \right) \right| d\mu_{\mathcal{X}} \\ &\leq \left\| \phi^{(L)} \right\|_{C^0}^{z-(m+n-1)} \int_{\mathcal{X}} \left| \left( \phi^{(L)} \circ a^{s_{i_1}} \right) \cdots \left( \phi^{(L)} \circ a^{s_{i_{m+n-1}}} \right) \right| d\mu_{\mathcal{X}} \\ &\ll_f \left\| \hat{f}^{(L)} \right\|_{C^0}^{z-(m+n-1)} \left\| \hat{f}^{(L)} \right\|_{L^{m+n-1}(\mathcal{X})}^{m+n-1}. \end{aligned}$$

This implies (5.18) and (5.17).

Finally, we recall that if  $(s_1, \dots, s_r) \in \Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; M, N)$  with  $\mathcal{Q} = \{\{0\}, \{1, \dots, r\}\}$ , then we have  $|s_{i_1} - s_{i_2}| \leq \alpha_j$  for all  $i_1 \neq i_2$ , and thus

$$|\Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; M, N)| \ll N \alpha_j^{r-1}. \quad (5.19)$$

Combining (5.17) and (5.19), we conclude that

$$\begin{aligned} &\frac{1}{N^{r/2}} \sum_{(s_1, \dots, s_r) \in \Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; M, N)} \left| \text{Cum}_{\mu_{\mathcal{X}}}^{(r)} \left( \phi^{(L)} \circ a^{s_1}, \dots, \phi^{(L)} \circ a^{s_r} \right) \right| \\ &\ll_f N^{1-r/2} \alpha_j^{r-1} \left\| \hat{f}^{(L)} \right\|_{C^0}^{(r-(m+n-1))^+} \left\| \hat{f}^{(L)} \right\|_{L^{m+n-1}(\mathcal{X})}^{\min(r, m+n-1)}. \end{aligned}$$

Hence, it follows from (5.16) that

$$\begin{aligned} &\frac{1}{N^{r/2}} \sum_{(s_1, \dots, s_r) \in \Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; M, N)} \text{Cum}_{\mu_{\mathcal{Y}}}^{(r)} \left( \psi_{s_1}^{(L)}, \dots, \psi_{s_r}^{(L)} \right) \\ &\ll_f N^{r/2} e^{-(\delta\beta_{j+1} - r\alpha_j\xi)} \left\| \hat{f}^{(L)} \right\|_{C^k}^r \\ &\quad + N^{1-r/2} \alpha_j^{r-1} \left\| \hat{f}^{(L)} \right\|_{C^0}^{(r-(m+n-1))^+} \left\| \hat{f}^{(L)} \right\|_{L^{m+n-1}(\mathcal{X})}^{\min(r, m+n-1)} \end{aligned}$$

and using Lemma 4.12, we deduce that

$$\begin{aligned} &\frac{1}{N^{r/2}} \sum_{(s_1, \dots, s_r) \in \Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; M, N)} \text{Cum}_{\mu_{\mathcal{Y}}}^{(r)} \left( \psi_{s_1}^{(L)}, \dots, \psi_{s_r}^{(L)} \right) \\ &\ll_f N^{r/2} e^{-(\delta\beta_{j+1} - r\alpha_j\xi)} L^r \|f\|_{C^k} + N^{1-r/2} \alpha_j^{r-1} L^{(r-(m+n-1))^+} \|f\|_{C^0}^r. \end{aligned}$$

### 5.1.2 Case 2: Summing over $(s_1, \dots, s_r) \in \Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; M, N)$ with $|\mathcal{Q}| \geq 2$ and $\mathcal{Q} \neq \{0\}, \{1, \dots, r\}$

In this case, the estimate (3.24) gives for  $(s_1, \dots, s_r) \in \Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; M, N)$ ,

$$\left| \text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(\psi_{s_1}^{(L)}, \dots, \psi_{s_r}^{(L)}) \right| \ll e^{-(\delta\beta_{j+1}-r\xi\alpha_j)} \left\| \hat{f}^{(L)} \right\|_{C^k}^r,$$

and

$$\begin{aligned} \sum_{(s_1, \dots, s_r) \in \Omega_{\mathcal{Q}}(\alpha_j, \beta_{j+1}; M, N)} \left| \text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(\psi_{s_1}^{(L)}, \dots, \psi_{s_r}^{(L)}) \right| &\ll N^{r/2} e^{-(\delta\beta_{j+1}-r\xi\alpha_j)} \left\| \hat{f}^{(L)} \right\|_{C^k}^r \\ &\ll N^{r/2} e^{-(\delta\beta_{j+1}-r\xi\alpha_j)} L^r \|f\|_{C^k}^r, \end{aligned}$$

where we used Lemma 4.12.

Finally, we combine the established bounds to estimate  $\text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(F_N^{(L)})$ . We choose the parameters  $\alpha_j$  and  $\beta_j$  as in (3.27). Then  $\beta_{r+1} \ll_r \gamma$ . In particular, we may choose

$$M \gg_r \gamma \quad (5.20)$$

to guarantee that (5.10) is satisfied. With these choices of  $\alpha_j$  and  $\beta_j$ , we obtain the estimate

$$\left| \text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(F_N^{(L)}) \right| \ll_f N^{r/2} e^{-\delta\gamma} L^r \|f\|_{C^k}^r + N^{1-r/2} \gamma^{r-1} L^{(r-(m+n-1))^+} \|f\|_{C^0}^r. \quad (5.21)$$

We observe that since  $m \geq 2$ ,

$$\frac{(r - (m + n - 1))^+}{m + n} < r/2 - 1 \quad \text{for all } r \geq 3,$$

Hence, we can choose  $q > 1/(m + n)$  such that

$$q(r - (m + n - 1))^+ < r/2 - 1 \quad \text{for all } r \geq 3.$$

Then we select

$$L = N^q,$$

so that, in particular, the condition (5.7) is satisfied. Now (5.21) can be rewritten as

$$\left| \text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(F_N^{(L)}) \right| \ll_f N^{r/2+rq} e^{-\delta\gamma} \|f\|_{C^k}^r + N^{q(r-(m+n-1))^+-(r/2-1)} \gamma^{r-1} \|f\|_{C^0}^r. \quad (5.22)$$



Choosing  $\gamma$  of the form

$$\gamma = c_r(\log N)$$

with sufficiently large  $c_r > 0$ , we conclude that

$$\text{Cum}_{\mu_Y}^{(r)}(F_N^{(L)}) \rightarrow 0 \quad \text{as } N \rightarrow \infty$$

for all  $r \geq 3$ .

## 5.2 Estimating variances

We now turn to the analysis of the variances of the average  $F_N^{(L)}$  which are given by

$$\|F_N^{(L)}\|_{L^2(\mathcal{Y})}^2 = \frac{1}{N} \sum_{s_1=M}^{N-1} \sum_{s_2=M}^{N-1} \int_{\mathcal{Y}} \psi_{s_1}^{(L)} \psi_{s_2}^{(L)} d\mu_Y.$$

We proceed as in Sect. 3.3 taking into account dependence on parameters  $M$  and  $L$ . We observe that this expression is symmetric with respect to  $s_1$  and  $s_2$ , writing  $s_1 = s + t$  and  $s_2 = t$  with  $0 \leq s \leq N - M - 1$  and  $M \leq t \leq N - s - 1$ , we obtain that

$$\|F_N^{(L)}\|_{L^2(\mathcal{Y})}^2 = \Theta_N^{(L)}(0) + 2 \sum_{s=1}^{N-M-1} \Theta_N^{(L)}(s), \quad (5.23)$$

where

$$\Theta_N^{(L)}(s) := \frac{1}{N} \sum_{t=M}^{N-1-s} \int_{\mathcal{Y}} \psi_{s+t}^{(L)} \psi_t^{(L)} d\mu_Y.$$

We have

$$\int_{\mathcal{Y}} \psi_{s+t}^{(L)} \psi_t^{(L)} d\mu_Y = \int_{\mathcal{Y}} (\hat{f}^{(L)} \circ a^{s+t})(\hat{f}^{(L)} \circ a^t) d\mu_Y - \mu_Y(\hat{f}^{(L)} \circ a^{s+t}) \mu_Y(\hat{f}^{(L)} \circ a^t).$$

To estimate  $\Theta_N^{(L)}(s)$ , we introduce an additional parameter  $K = K(N) \rightarrow \infty$  such that  $K \leq M$  (to be specified later) and consider separately the cases when  $s < K$  and when  $s \geq K$ .

First, we consider the case when  $s \geq K$ . By Corollary 3.3, we have

$$\int_{\mathcal{Y}} (\hat{f}^{(L)} \circ a^{s+t})(\hat{f}^{(L)} \circ a^t) d\mu_Y = \mu_{\mathcal{X}}(\hat{f}^{(L)})^2 + O\left(e^{-\delta \min(s,t)} \|\hat{f}^{(L)}\|_{C^k}^2\right). \quad (5.24)$$

Also, by Corollary 3.3,

$$\int_{\mathcal{Y}} (\hat{f}^{(L)} \circ a^t) d\mu_Y = \mu_{\mathcal{X}}(\hat{f}^{(L)}) + O\left(e^{-\delta t} \|\hat{f}^{(L)}\|_{C^k}\right). \quad (5.25)$$

Hence, combining (5.24) and (5.25), we deduce that

$$\int_{\mathcal{Y}} \psi_{s+t}^{(L)} \psi_t^{(L)} d\mu_{\mathcal{Y}} = O\left(e^{-\delta \min(s,t)} \|\hat{f}^{(L)}\|_{C^k}^2\right).$$

Since

$$\sum_{s=K}^{N-M-1} \left( \sum_{t=M}^{N-1-s} e^{-\delta \min(s,t)} \right) \leq \sum_{s=K}^{N-1} \sum_{t=M}^{N-1} (e^{-\delta s} + e^{-\delta t}) \ll N e^{-\delta K},$$

we conclude that

$$\sum_{s=K}^{N-M-1} \Theta_N^{(L)}(s) \ll e^{-\delta K} \|\hat{f}^{(L)}\|_{C^k}^2 \ll_f e^{-\delta K} L^2 \|f\|_{C^k}^2, \quad (5.26)$$

where we used Lemma 4.12. The implied constants here and below in the proof depend only on  $\text{supp}(f)$ .

Let us now consider the case  $s < K$ . We observe that Corollary 3.3 (for  $r = 1$ ) applied to the function  $\phi_s^{(L)} := (\hat{f}^{(L)} \circ a^s) \hat{f}^{(L)}$  yields,

$$\begin{aligned} \int_{\mathcal{Y}} (\hat{f}^{(L)} \circ a^{s+t})(\hat{f}^{(L)} \circ a^t) d\mu_{\mathcal{Y}} &= \int_{\mathcal{Y}} (\phi_s^{(L)} \circ a^t) d\mu_{\mathcal{Y}} \\ &= \int_{\mathcal{X}} \phi_s^{(L)} d\mu_{\mathcal{X}} + O\left(e^{-\delta t} \|\phi_s^{(L)}\|_{C^k}\right). \end{aligned}$$

Furthermore, for some  $\xi = \xi(m, n, k) > 0$ , we have

$$\|\phi_s^{(L)}\|_{C^k} \ll \|\hat{f}^{(L)} \circ a^s\|_{C^k} \|\hat{f}^{(L)}\|_{C^k} \ll e^{\xi s} \|\hat{f}^{(L)}\|_{C^k}^2.$$

Therefore, we deduce that

$$\begin{aligned} \int_{\mathcal{Y}} (\hat{f}^{(L)} \circ a^{s+t})(\hat{f}^{(L)} \circ a^t) d\mu_{\mathcal{Y}} &= \int_{\mathcal{X}} (\hat{f}^{(L)} \circ a^s) \hat{f}^{(L)} d\mu_{\mathcal{X}} \\ &\quad + O\left(e^{-\delta t} e^{\xi s} \|\hat{f}^{(L)}\|_{C^k}^2\right). \end{aligned}$$

Combining this estimate with (5.25), we conclude that

$$\begin{aligned} \int_{\mathcal{Y}} \psi_{s+t}^{(L)} \psi_t^{(L)} d\mu_{\mathcal{Y}} &= \int_{\mathcal{X}} (\hat{f}^{(L)} \circ a^s) \hat{f}^{(L)} d\mu_{\mathcal{X}} - \mu_{\mathcal{X}}(\hat{f}^{(L)})^2 \\ &\quad + O\left(e^{-\delta t} e^{\xi s} \|\hat{f}^{(L)}\|_{C^k}^2\right). \end{aligned}$$

Hence, setting

$$\Theta_{\infty}^{(L)}(s) := \int_{\mathcal{X}} (\hat{f}^{(L)} \circ a^s) \hat{f}^{(L)} d\mu_{\mathcal{X}} - \mu_{\mathcal{X}}(\hat{f}^{(L)})^2,$$

we obtain that

$$\begin{aligned} \Theta_N^{(L)}(s) &= \frac{N-M-s}{N} \Theta_{\infty}^{(L)}(s) + O\left(N^{-1} e^{-\delta M} e^{\xi s} \left\| \hat{f}^{(L)} \right\|_{C^k}^2\right) \\ &= \Theta_{\infty}^{(L)}(s) + O\left(N^{-1}(M+s) \left\| \hat{f}^{(L)} \right\|_{L^2(\mathcal{X})}^2 + N^{-1} e^{-\delta M} e^{\xi s} \left\| \hat{f}^{(L)} \right\|_{C^k}^2\right). \end{aligned}$$

Therefore, using Lemma 4.12, we deduce that

$$\begin{aligned} \Theta_N^{(L)}(0) + 2 \sum_{s=1}^{K-1} \Theta_N^{(L)}(s) &= \Theta_{\infty}^{(L)}(0) + 2 \sum_{s=1}^{K-1} \Theta_{\infty}^{(L)}(s) \\ &\quad + O_f\left(N^{-1}(M+K)K \|f\|_{C^0}^2 + N^{-1} e^{-\delta M} e^{\xi K} L^2 \|f\|_{C^k}^2\right). \end{aligned} \quad (5.27)$$

Combining (5.26) and (5.27), we conclude that

$$\begin{aligned} \Theta_N^{(L)}(0) + 2 \sum_{s=1}^{M-N-1} \Theta_N^{(L)}(s) \\ &= \Theta_{\infty}^{(L)}(0) + 2 \sum_{s=1}^{K-1} \Theta_{\infty}^{(L)}(s) \\ &\quad + O_f\left(N^{-1}(M+K)K \|f\|_{C^0}^2 + (N^{-1} e^{-\delta M} e^{\xi K} + e^{-\delta K}) L^2 \|f\|_{C^k}^2\right). \end{aligned} \quad (5.28)$$

We choose the parameters  $K = K(N)$ ,  $M = M(N)$ , and  $L = L(N)$  so that

$$e^{-\delta K} L^2 \rightarrow 0, \quad (5.29)$$

$$N^{-1} e^{-\delta M} e^{\xi K} L^2 \rightarrow 0, \quad (5.30)$$

$$N^{-1}(M+K)K \rightarrow 0 \quad (5.31)$$

as  $N \rightarrow \infty$ . Then

$$\left\| F_N^{(L)} \right\|_{L^2(\mathcal{Y})}^2 = \Theta_{\infty}^{(L)}(0) + 2 \sum_{s=1}^{K-1} \Theta_{\infty}^{(L)}(s) + o(1).$$

Next, we shall show that with a suitable choice of parameters,

$$\left\| F_N^{(L)} \right\|_{L^2(\mathcal{Y})}^2 = \Theta_{\infty}(0) + 2 \sum_{s=1}^{K-1} \Theta_{\infty}(s) + o(1), \quad (5.32)$$

where

$$\Theta_{\infty}(s) := \int_{\mathcal{X}} (\hat{f} \circ a^s) \hat{f} d\mu_{\mathcal{X}} - \mu_{\mathcal{X}}(\hat{f})^2.$$

We recall that by Lemma 4.12, for all  $\tau < m + n - 1$ ,

$$\|\hat{f} - \hat{f}^{(L)}\|_{L^1(\mathcal{X})} \ll_{f,\tau} L^{-\tau} \|f\|_{C^0} \quad \text{and} \quad \|\hat{f} - \hat{f}^{(L)}\|_{L^2(\mathcal{X})} \ll_{f,\tau} L^{-(\tau-1)/2} \|f\|_{C^0}, \quad (5.33)$$

where the implied constant depends only on  $\text{supp}(f)$ . It follows from these estimates that

$$\begin{aligned} \mu_{\mathcal{X}}(\hat{f}^{(L)}) &= \mu_{\mathcal{X}}(\hat{f}) + O_{f,\tau}(L^{-\tau} \|f\|_{C^0}), \\ \int_{\mathcal{X}} (\hat{f}^{(L)} \circ a^s) \hat{f}^{(L)} d\mu_{\mathcal{X}} &= \int_{\mathcal{X}} (\hat{f} \circ a^s) \hat{f} d\mu_{\mathcal{X}} + O_{f,\tau}(L^{-(\tau-1)/2} \|f\|_{C^0}^2), \end{aligned}$$

so that

$$\Theta_{\infty}^{(L)}(s) = \Theta_{\infty}(s) + O_{f,\tau}(L^{-(\tau-1)/2} \|f\|_{C^0}^2). \quad (5.34)$$

We choose the parameters  $K = K(N) \rightarrow \infty$  and  $L = L(N) \rightarrow \infty$  so that

$$KL^{-(\tau-1)/2} \rightarrow 0 \quad \text{for some } \tau < m + n - 1. \quad (5.35)$$

Then (5.32) follows. We conclude that

$$\|F_N^{(L)}\|_{L^2(\mathcal{Y})}^2 \rightarrow \Theta_{\infty}(0) + 2 \sum_{s=1}^{\infty} \Theta_{\infty}(s) \quad (5.36)$$

as  $N \rightarrow \infty$ .

Finally, we compute  $\Theta_{\infty}(s)$  by using Rogers formula (Proposition 4.4) applied to the function

$$F_s(\bar{z}_1, \bar{z}_2) := \sum_{p,q \geq 1} f(pa^s \bar{z}_1) f(q\bar{z}_2), \quad (\bar{z}_1, \bar{z}_2) \in \mathbb{R}^{m+n} \times \mathbb{R}^{m+n}.$$

Since

$$\int_{\mathcal{X}} (\hat{f} \circ a^s) \hat{f} d\mu_{\mathcal{X}} = \int_{\mathcal{X}} \left( \sum_{(\bar{z}_1, \bar{z}_2) \in (P(\mathbb{Z}^{m+n}))} F_s(g\bar{z}_1, g\bar{z}_2) \right) d\mu_{\mathcal{X}}(g\Gamma),$$

we deduce that

$$\int_{\mathcal{X}} (\hat{f} \circ a^s) \hat{f} d\mu_{\mathcal{X}} = \left( \int_{\mathbb{R}^{m+n}} f(\bar{z}) d\bar{z} \right)^2 \\ + \zeta(m+n)^{-1} \sum_{p,q \geq 1} \left( \int_{\mathbb{R}^{m+n}} f(pa^s \bar{z}) f(q\bar{z}) d\bar{z} + \int_{\mathbb{R}^{m+n}} f(pa^s \bar{z}) f(-q\bar{z}) d\bar{z} \right).$$

Since by the Siegel Mean Value Theorem (Proposition 4.3),

$$\int_{\mathcal{X}} \hat{f} d\mu_{\mathcal{X}} = \int_{\mathbb{R}^{m+n}} f(\bar{z}) d\bar{z},$$

we conclude that

$$\Theta_{\infty}(s) = \zeta(m+n)^{-1} \sum_{p,q \geq 1} \left( \int_{\mathbb{R}^{m+n}} f(pa^s \bar{z}) f(q\bar{z}) dz + \int_{\mathbb{R}^{m+n}} f(pa^s \bar{z}) f(-q\bar{z}) d\bar{z} \right).$$

Finally, we show that the sum in (5.36) is finite. We represent points  $\bar{z} \in \mathbb{R}^{m+n}$  as  $\bar{z} = (\bar{x}, \bar{y})$  with  $\bar{x} \in \mathbb{R}^m$  and  $\bar{y} \in \mathbb{R}^n$ . Since  $f$  is bounded, and its compact support is contained in  $\{\bar{y} \neq 0\}$ , we may assume without loss of generality that  $f$  is the characteristic function of the set

$$\{(\bar{x}, \bar{y}) \in \mathbb{R}^{m+n} : \nu_1 \leq \|\bar{y}\| \leq \nu_2, \quad |x_i| \leq \vartheta \|\bar{y}\|^{-w_i}, \quad i = 1, \dots, m\}$$

with  $0 < \nu_1 < \nu_2$  and  $\vartheta > 0$ . Let

$$\Omega_s(p) := \left\{ (\bar{x}, \bar{y}) \in \mathbb{R}^{m+n} : \frac{\nu_1 e^s}{p} \leq \|\bar{y}\| \leq \frac{\nu_2 e^s}{p}, \quad p^{1+w_i} |x_i| \|\bar{y}\|^{w_i} \leq \vartheta, \quad i = 1, \dots, m \right\}.$$

Then

$$\int_{\mathbb{R}^{m+n}} f(pa^s \bar{z}) f(\pm q\bar{z}) d\bar{z} = \text{vol}(\Omega_s(p) \cap \Omega_0(q)).$$

Setting

$$I(u) := \{\bar{y} \in \mathbb{R}^n : \nu_1 u \leq \|\bar{y}\| \leq \nu_2 u\},$$

we obtain that

$$\text{vol}(\Omega_s(p) \cap \Omega_0(q)) = \int_{I(e^s p^{-1}) \cap I(q^{-1})} \left( \prod_{i=1}^m 2\vartheta \max(p, q)^{-1-w_i} \|\bar{y}\|^{-w_i} \right) d\bar{y} \\ \ll \max(p, q)^{-m-n} \int_{I(e^s p^{-1}) \cap I(q^{-1})} \|\bar{y}\|^{-n} d\bar{y}.$$

We note that  $I(e^s p^{-1}) \cap I(q^{-1}) = \emptyset$  unless  $(v_1 v_2^{-1})e^s q \leq p \leq (v_2 v_1^{-1})e^s q$ , and also that

$$\int_{I(q^{-1})} \|\bar{y}\|^{-n} d\bar{y} \ll \int_{v_1/q}^{v_2/q} r^{-1} dr \ll 1.$$

Hence, it follows that

$$\begin{aligned} \sum_{s=1}^{\infty} \Theta_{\infty}(s) &\ll \sum_{s,q=1}^{\infty} \sum_{(v_1 v_2^{-1})e^s q \leq p \leq (v_2 v_1^{-1})e^s q} \max(p, q)^{-(m+n)} \\ &\ll \sum_{s,q=1}^{\infty} (e^s q)^{-(m+n-1)} < \infty, \end{aligned}$$

because  $m + n \geq 3$ .

### 5.3 Proof of Theorem 5.1

As we already remarked above, it is sufficient to show that the sequence of averages  $F_N^{(L)}$  converges in distribution to the normal law. To verify this, we use the Method of Cumulants (Proposition 3.4). It is easy to see that

$$\int_{\mathcal{Y}} F_N^{(L)} d\mu_{\mathcal{Y}} = 0.$$

Moreover, with a suitable choice of parameters, we have shown in Sect. 5.1 that for  $r \geq 3$ ,

$$\text{Cum}_{\mu_{\mathcal{Y}}}^{(r)}(F_N^{(L)}) \rightarrow 0 \quad \text{as } N \rightarrow \infty,$$

and in Sect. 5.2 that

$$\|F_N^{(L)}\|_{L^2(\mathcal{Y})}^2 \rightarrow \sigma_f^2 < \infty \quad \text{as } N \rightarrow \infty.$$

Hence, Proposition 3.4 applies, and it remains to verify that we can choose our parameters that satisfy the stated assumptions. We recall that

$$L = N^q \quad \text{and} \quad \gamma = c_r(\log N).$$

The parameters  $M = M(N) \geq K = K(N)$  need to satisfy the seven conditions (5.4), (5.5), (5.20), (5.29), (5.30), (5.31), (5.35). We take

$$K(N) = c_1(\log N)$$

with sufficiently large  $c_1 > 0$  so that (5.29) is satisfied. Then taking

$$M(N) = (\log N)(\log \log N),$$

we arrange that (5.5), (5.20), and (5.30) hold for all  $N \geq N_0(r)$ . We note that the constant  $c_r$  and the implicit constant in (5.20) depends on  $r$ , and the  $(\log \log N)$ -factor here is added to guarantee that the parameter  $M$  is independent of  $r$ . Finally, the conditions (5.4), (5.31), (5.35) are immediate from our choices.

## 6 CLT for counting functions and the proof of Theorem 1.2

We recall from Sect. 1.3 that

$$\Delta_T(u) = |\Lambda_u \cap \Omega_T| + O(1),$$

where  $\Lambda_u$  is defined in (1.8) and the domains  $\Omega_T$  are defined in (1.9). We shall decompose this domain into smaller pieces using the linear map  $a = \text{diag}(e^{w_1}, \dots, e^{w_m}, e^{-1}, \dots, e^{-1})$ . We note that for any integer  $N \geq 1$ ,

$$\Omega_{e^N} = \bigsqcup_{s=0}^{N-1} a^{-s} \Omega_e,$$

and thus

$$|\Lambda_u \cap \Omega_{e^N}| = \sum_{s=0}^{N-1} \hat{\chi}(a^s \Lambda_u),$$

where  $\chi$  denotes the characteristic function of the set  $\Omega_e$ . Hence the proof of Theorem 1.2 reduces to analyzing sums of the form  $\sum_{s=0}^{N-1} \hat{\chi}(a^s y)$  with  $y \in \mathcal{Y}$ . For this purpose, we define

$$F_N := \frac{1}{\sqrt{N}} \sum_{s=0}^{N-1} (\hat{\chi} \circ a^s - \mu_{\mathcal{Y}}(\hat{\chi} \circ a^s)). \quad (6.1)$$

Our main result in this section now reads as follows.

**Theorem 6.1** *If  $m \geq 2$ , then for every  $\xi \in \mathbb{R}$ ,*

$$\mu_{\mathcal{Y}}(\{y \in \mathcal{Y} : F_N(y) < \xi\}) \rightarrow \text{Norm}_{\sigma}(\xi)$$

as  $N \rightarrow \infty$ , where

$$\sigma^2 := 2^{m+1} \left( \prod_{i=1}^m \vartheta_i \right) \left( \int_{S^{n-1}} \|\bar{z}\|^{-n} d\bar{z} \right) \left( \frac{2\zeta(m+n-1)}{\zeta(m+n)} - 1 \right).$$

We approximate  $\chi$  by a family of non-negative functions  $f_\varepsilon \in C_c^\infty(\mathbb{R}^{m+n})$  whose supports are contained in an  $\varepsilon$ -neighbourhood of the set  $\Omega_e$ , and

$$\chi \leq f_\varepsilon \leq 1, \quad \|f_\varepsilon - \chi\|_{L^1(\mathbb{R}^{m+n})} \ll \varepsilon, \quad \|f_\varepsilon - \chi\|_{L^2(\mathbb{R}^{m+n})} \ll \varepsilon^{1/2}, \quad \|f_\varepsilon\|_{C^k} \ll \varepsilon^{-k}.$$

This approximation allows us to construct smooth approximations of the Siegel transform  $\hat{\chi}$  in the following sense.

**Proposition 6.2** *For every  $s \geq 0$ ,*

$$\int_{\mathcal{Y}} \left| \hat{f}_\varepsilon \circ a^s - \hat{\chi} \circ a^s \right| d\mu_{\mathcal{Y}} \ll \varepsilon + e^{-s}.$$

**Proof** We observe that there exists  $\vartheta_i(\varepsilon) > \vartheta_i$  such that  $\vartheta_i(\varepsilon) = \vartheta_i + O(\varepsilon)$  and  $f_\varepsilon \leq \chi_\varepsilon$ , where  $\chi_\varepsilon$  denotes the characteristic function of the set

$$\{(\bar{x}, \bar{y}) \in \mathbb{R}^{m+n} : 1 - \varepsilon \leq \|\bar{y}\| \leq e + \varepsilon, |x_i| < \vartheta_i(\varepsilon) \|\bar{y}\|^{-w_i} \text{ for } i = 1, \dots, m\}.$$

Then it follows that

$$|\hat{f}_\varepsilon(a^s \Lambda) - \hat{\chi}(a^s \Lambda)| = \sum_{v \in \Lambda \setminus \{0\}} (f_\varepsilon(a^s v) - \chi(a^s v)) \leq \sum_{v \in \Lambda \setminus \{0\}} (\chi_\varepsilon(a^s v) - \chi(a^s v)).$$

It is clear that  $\chi_\varepsilon - \chi$  is bounded by the sum  $\chi_{1,\varepsilon} + \chi_{2,\varepsilon} + \chi_{3,\varepsilon}$  of the characteristic functions of the sets

$$\begin{aligned} & \{(\bar{x}, \bar{y}) \in \mathbb{R}^{m+n} : 1 - \varepsilon \leq \|\bar{y}\| \leq 1, |x_i| < \vartheta_i(\varepsilon) \|\bar{y}\|^{-w_i} \text{ for } i = 1, \dots, m\}, \\ & \{(\bar{x}, \bar{y}) \in \mathbb{R}^{m+n} : e \leq \|\bar{y}\| \leq e + \varepsilon, |x_i| < \vartheta_i(\varepsilon) \|\bar{y}\|^{-w_i} \text{ for } i = 1, \dots, m\}, \\ & \{(\bar{x}, \bar{y}) \in \mathbb{R}^{m+n} : 1 \leq \|\bar{y}\| \leq e, |x_i| < \vartheta_i(\varepsilon) \|\bar{y}\|^{-w_i} \\ & \quad \text{for all } i, |x_j| \geq \vartheta_j \|\bar{y}\|^{-w_j} \text{ for some } j\} \end{aligned}$$

respectively. In particular, we obtain that

$$\hat{f}_\varepsilon(a^s \Lambda) - \hat{\chi}(a^s \Lambda) \leq \hat{\chi}_{1,\varepsilon}(a^s \Lambda) + \hat{\chi}_{2,\varepsilon}(a^s \Lambda) + \hat{\chi}_{3,\varepsilon}(a^s \Lambda).$$

Hence, it remains to show that for  $j = 1, 2, 3$ ,

$$\int_{\mathcal{Y}} (\hat{\chi}_{j,\varepsilon} \circ a^s) d\mu_{\mathcal{Y}} \ll \varepsilon + e^{-s}.$$

As in (4.3), we compute that

$$\int_{\mathcal{Y}} (\hat{\chi}_{1,\varepsilon} \circ a^s) d\mu_{\mathcal{Y}} = \sum_{(1-\varepsilon)e^s \leq \|\bar{q}\| \leq e^s} \prod_{i=1}^m \left( \sum_{p_i \in \mathbb{Z}} \int_{[0,1]^n} \chi_{\vartheta_i(\varepsilon)\|\bar{q}\|^{-w_i}}(p_i + \langle \bar{u}_i, \bar{q} \rangle) d\bar{u}_i \right), \quad (6.2)$$



where  $\chi_\theta$  denotes the characteristic function of the interval  $[-\theta, \theta]$ . We observe that

$$\int_{[0,1]^n} \chi_{\vartheta_i(\varepsilon)\|\bar{q}\|^{-w_i}}(p_i + \langle \bar{u}_i, \bar{q} \rangle) d\bar{u}_i \ll (\max_k |q_k|)^{-1} \|\bar{q}\|^{-w_i} \ll \|\bar{q}\|^{-1-w_i},$$

and moreover this integral is non-zero only when  $|p_i| = O(\|\bar{q}\|)$ . Hence,

$$\sum_{p_i \in \mathbb{Z}} \int_{[0,1]^n} \chi_{\vartheta_i(\varepsilon)\|\bar{q}\|^{-w_i}}(p_i + \langle \bar{u}_i, \bar{q} \rangle) d\bar{u}_i \ll \|\bar{q}\|^{-w_i},$$

and

$$\begin{aligned} \int_{\mathcal{Y}} (\hat{\chi}_{1,\varepsilon} \circ a^s) d\mu_{\mathcal{Y}} &\ll \sum_{(1-\varepsilon)e^s \leq \|\bar{q}\| \leq e^s} \prod_{i=1}^m \|\bar{q}\|^{-w_i} \\ &\ll e^{-ns} \left| \{ \bar{q} \in \mathbb{Z}^n : (1-\varepsilon)e^s \leq \|\bar{q}\| \leq e^s \} \right|. \end{aligned}$$

The number of integral points in the region  $\{(1-\varepsilon)e^s \leq \|\bar{y}\| \leq e^s\}$  can be estimated in terms of its volume. Namely, there exist  $r > 0$  (depending only on the norm) such that

$$\left| \{ \bar{q} \in \mathbb{Z}^n : (1-\varepsilon)e^s \leq \|\bar{q}\| \leq e^s \} \right| \ll \left| \{ \bar{y} \in \mathbb{R}^n : (1-\varepsilon)e^s - r \leq \|\bar{y}\| \leq e^s + r \} \right|.$$

Hence,

$$\begin{aligned} \int_{\mathcal{Y}} (\hat{\chi}_{1,\varepsilon} \circ a^s) d\mu_{\mathcal{Y}} &\ll e^{-ns} ((e^s + r)^n - ((1-\varepsilon)e^s - r)^n) \\ &\ll (1 + re^{-s})^n - (1 - \varepsilon - re^{-s})^n \\ &\ll \varepsilon + e^{-s}. \end{aligned}$$

The integral for  $\hat{\chi}_{2,\varepsilon} \circ a^s$  can be estimated similarly.

The integral over  $\hat{\chi}_{3,\varepsilon} \circ a^s$  as in (6.2) can be written as a sum of the products of the integral

$$\begin{aligned} &\int_{[0,1]^n} \left( \chi_{\vartheta_j(\varepsilon)\|\bar{q}\|^{-w_j}}(p_j + \langle \bar{u}_j, \bar{q} \rangle) - \chi_{\vartheta_j\|\bar{q}\|^{-w_j}}(p_j + \langle \bar{u}_j, \bar{q} \rangle) \right) d\bar{u}_j \\ &\leq 2(\max_k |q_k|)^{-1} (\vartheta_j(\varepsilon) - \vartheta_j) \|\bar{q}\|^{-w_j} \ll \varepsilon \|\bar{q}\|^{-1-w_j}, \end{aligned}$$

and the integrals

$$\int_{[0,1]^n} \chi_{\vartheta_i(\varepsilon)\|\bar{q}\|^{-w_i}}(p_i + \langle \bar{u}_i, \bar{q} \rangle) d\bar{u}_i \leq 2\vartheta_i(\varepsilon)(\max_k |q_k|)^{-1} \|\bar{q}\|^{-w_i} \ll \|\bar{q}\|^{-1-w_i}$$

with  $i \neq j$ . We observe that these integrals are non-zero only when  $|p_j| = O(\|\bar{q}\|)$  and  $|p_i| = O(\|\bar{q}\|)$ . Hence, we conclude that

$$\int_{\mathcal{Y}} (\hat{\chi}_{3,\varepsilon} \circ a^s) d\mu_{\mathcal{Y}} \ll \sum_{e^s \leq \|\bar{q}\| \leq e^{s+1}} \varepsilon \prod_{i=1}^m \|\bar{q}\|^{-w_i} = \varepsilon \sum_{e^s \leq \|\bar{q}\| \leq e^{s+1}} \|\bar{q}\|^{-n} \ll \varepsilon,$$

which completes the proof of the proposition.  $\square$

Now we start with the proof of Theorem 6.1. As in Sect. 5, we modify  $F_N$  and consider instead

$$\tilde{F}_N := \frac{1}{\sqrt{N}} \sum_{s=M}^{N-1} (\hat{\chi} \circ a^s - \mu_{\mathcal{Y}}(\hat{\chi} \circ a^s)) \quad (6.3)$$

for a parameter  $M = M(N) \rightarrow \infty$  that will be chosen later. As in (5.3) we obtain that

$$\|F_N - \tilde{F}_N\|_{L^1(\mathcal{Y})} \rightarrow 0 \quad \text{as } N \rightarrow \infty$$

provided that

$$M = o(N^{1/2}). \quad (6.4)$$

Hence, if we can prove the CLT for  $(\tilde{F}_N)$ , then the CLT for  $(F_N)$  would follow. From now on, to simplify notations, we assume that  $F_N$  is given by (6.3).

Our next step is to exploit the approximation  $\chi \approx f_\varepsilon$ , so we introduce

$$F_N^{(\varepsilon)} := \frac{1}{\sqrt{N}} \sum_{s=M}^{N-1} (\hat{f}_\varepsilon \circ a^s - \mu_{\mathcal{Y}}(\hat{f}_\varepsilon \circ a^s)),$$

where the parameter  $\varepsilon = \varepsilon(N) \rightarrow 0$  will be specified later. We observe that it follows from Proposition 6.2 that

$$\|F_N^{(\varepsilon)} - F_N\|_{L^1(\mathcal{Y})} \leq \frac{2}{\sqrt{N}} \sum_{s=M}^{N-1} \|\hat{f}_\varepsilon \circ a^s - \hat{\chi} \circ a^s\|_{L^1(\mathcal{Y})} \ll N^{1/2}(\varepsilon + e^{-M}).$$

We choose  $\varepsilon = \varepsilon(N)$  and  $M = M(N)$  so that

$$N^{1/2}\varepsilon \rightarrow 0 \quad \text{and} \quad N^{1/2}e^{-M} \rightarrow 0. \quad (6.5)$$

Then

$$\|F_N^{(\varepsilon)} - F_N\|_{L^1(\mathcal{Y})} \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

Hence, it remains to prove convergence in distribution for the sequence  $F_N^{(\varepsilon)}$ .

We observe that the sequence  $F_N^{(\varepsilon)}$  fits into the framework of Sect. 5. However, we need to take into account the dependence on the new parameter  $\varepsilon$  and refine the previous estimates. It will be important for our argument that the supports of the functions  $f_\varepsilon$  are uniformly bounded,  $\|f_\varepsilon\|_{C^0} \ll 1$ , and  $\|f_\varepsilon\|_{C^k} \ll \varepsilon^{-k}$ .

As in Sect. 5, we consider the truncation

$$F_N^{(\varepsilon, L)} := \frac{1}{\sqrt{N}} \sum_{s=M}^{N-1} \left( \hat{f}_\varepsilon^{(L)} \circ a^s - \mu_Y(\hat{f}_\varepsilon^{(L)} \circ a^s) \right).$$

defined for a parameter  $L = L(N) \rightarrow \infty$ . We assume that

$$M \gg \log L, \quad (6.6)$$

so that Proposition 4.8 applies when  $s \geq M$ . Since the family of functions  $f_\varepsilon$  is majorized by a fixed bounded function with compact support, Proposition 4.8 implies that when  $m \geq 2$ ,

$$\left\| \hat{f}_\varepsilon \circ a^s \right\|_{L^2(Y)} \ll 1 \quad \text{for all } s \geq 0,$$

uniformly on  $\varepsilon$ . Hence, the bound (5.6) can be proved exactly as before, and we obtain

$$\left\| F_N^{(\varepsilon)} - F_N^{(\varepsilon, L)} \right\|_{L^1(Y)} \ll_p N^{1/2} L^{-p/2} \quad \text{for all } p < m + n.$$

We choose the parameter  $L$  as before so that

$$N = o(L^p) \quad \text{for some } p < m + n \quad (6.7)$$

to guarantee that

$$\left\| F_N^{(\varepsilon)} - F_N^{(\varepsilon, L)} \right\|_{L^1(Y)} \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

Now it remains to show that the family  $F_N^{(\varepsilon, L)}$  satisfies the CLT with a suitable choice of parameters  $M, L, \varepsilon$ . As in Sect. 5 we will show that for  $r \geq 3$ ,

$$\text{Cum}_{\mu_Y}^{(r)} \left( F_N^{(\varepsilon, L)} \right) \rightarrow 0 \quad \text{as } N \rightarrow \infty, \quad (6.8)$$

and

$$\left\| F_N^{(\varepsilon, L)} \right\|_{L^2(Y)}^2 \rightarrow \sigma \quad \text{as } N \rightarrow \infty \quad (6.9)$$

with an explicit  $\sigma \in (0, \infty)$ .

Under the condition

$$M \gg \gamma, \quad (6.10)$$

the estimate (5.21) gives the bound

$$\begin{aligned} \left| \text{Cum}_{\mu_Y}^{(r)}(F_N^{(\varepsilon, L)}) \right| &\ll N^{r/2} e^{-\delta\gamma} L^r \|f_\varepsilon\|_{C^k}^r + N^{1-r/2} \gamma^{r-1} L^{(r-(m+n-1))^+} \|f_\varepsilon\|_{C^0}^r \\ &\ll N^{r/2} e^{-\delta\gamma} L^r \varepsilon^{-rk} + N^{1-r/2} \gamma^{r-1} L^{(r-(m+n-1))^+}. \end{aligned}$$

We note that the implicit constant in (5.21) depends only on  $\text{supp}(f_\varepsilon)$  so that it is uniform on  $\varepsilon$ . We choose  $L = N^q$  as in Sect. 5 and  $\gamma = c_r(\log N)$ , where  $c_r > 0$  will be specified later. In particular, then  $N^{1-r/2} \gamma^{r-1} L^{(r-(m+n-1))^+} \rightarrow 0$ , and assuming that

$$N^{r/2} L^r \varepsilon^{-rk} = o(e^{\delta\gamma}), \quad (6.11)$$

it follows that (6.8) holds.

To prove (6.9), we have to estimate

$$\left\| F_N^{(\varepsilon, L)} \right\|_{L^2(\mathcal{Y})}^2 = \frac{1}{N} \sum_{s_1=M}^{N-1} \sum_{s_2=M}^{N-1} \int_{\mathcal{Y}} \psi_{s_1}^{(\varepsilon, L)} \psi_{s_2}^{(\varepsilon, L)} d\mu_Y,$$

where

$$\psi_s^{(\varepsilon, L)}(y) := \hat{f}_\varepsilon^{(L)}(a^s y) - \mu_Y(\hat{f}_\varepsilon^{(L)} \circ a^s).$$

As in (5.23), we obtain that

$$\left\| F_N^{(\varepsilon, L)} \right\|_{L^2(\mathcal{Y})}^2 = \Theta_N^{(\varepsilon, L)}(0) + 2 \sum_{s=1}^{N-M-1} \Theta_N^{(\varepsilon, L)}(s),$$

where

$$\Theta_N^{(\varepsilon, L)}(s) := \frac{1}{N} \sum_{t=M}^{N-1-s} \int_{\mathcal{Y}} \psi_{s+t}^{(\varepsilon, L)} \psi_t^{(\varepsilon, L)} d\mu_Y.$$

Our estimate proceeds as in Sect. 5, and we shall show that with a suitable choice of parameters,

$$\left\| F_N^{(\varepsilon, L)} \right\|_{L^2(\mathcal{Y})}^2 = \Theta_\infty^{(\varepsilon, L)}(0) + 2 \sum_{s=1}^{K-1} \Theta_\infty^{(\varepsilon, L)}(s) + o(1), \quad (6.12)$$

where

$$\Theta_{\infty}^{(\varepsilon, L)}(s) := \int_{\mathcal{X}} (\hat{f}_{\varepsilon}^{(L)} \circ a^s) \hat{f}_{\varepsilon}^{(L)} d\mu_{\mathcal{X}} - \mu_{\mathcal{X}}(\hat{f}_{\varepsilon}^{(L)})^2.$$

Indeed, arguing as in (5.28), we deduce that

$$\begin{aligned} & \Theta_N^{(\varepsilon, L)}(0) + 2 \sum_{s=1}^{M-N-1} \Theta_N^{(\varepsilon, L)}(s) \\ &= \Theta_{\infty}^{(\varepsilon, L)}(0) + 2 \sum_{s=1}^{K-1} \Theta_{\infty}^{(\varepsilon, L)}(s) \\ & \quad + O\left(N^{-1}(M+K)K + N^{-1}e^{-\delta M}e^{\xi K}L^2\varepsilon^{-2k} + e^{-\delta K}L^2\varepsilon^{-2k}\right). \end{aligned}$$

Hence, (6.12) holds provided that

$$e^{-\delta K}L^2\varepsilon^{-2k} \rightarrow 0, \quad (6.13)$$

$$N^{-1}e^{-\delta M}e^{\xi K}L^2\varepsilon^{-2k} \rightarrow 0, \quad (6.14)$$

$$N^{-1}(M+K)K \rightarrow 0. \quad (6.15)$$

Next, we set

$$\Theta_{\infty}^{(\varepsilon)}(s) := \int_{\mathcal{X}} (\hat{f}_{\varepsilon} \circ a^s) \hat{f}_{\varepsilon} d\mu_{\mathcal{X}} - \mu_{\mathcal{X}}(\hat{f}_{\varepsilon})^2$$

and observe that as in (5.34),

$$\Theta_{\infty}^{(\varepsilon, L)}(s) = \Theta_{\infty}^{(\varepsilon)}(s) + O_{\tau}\left(L^{-(\tau-1)/2}\right) \quad \text{for all } \tau < m+n-1,$$

uniformly on  $\varepsilon$ . Hence, choosing  $K$  so that

$$KL^{-(\tau-1)/2} \rightarrow 0 \quad \text{for some } \tau < m+n-1, \quad (6.16)$$

we conclude that

$$\left\| F_N^{(\varepsilon, L)} \right\|_{L^2(\mathcal{Y})}^2 = \Theta_{\infty}^{(\varepsilon)}(0) + 2 \sum_{s=1}^{K-1} \Theta_{\infty}^{(\varepsilon)}(s) + o(1).$$

The terms  $\Theta^{(\varepsilon)}(s)$  can be computed using Propositions 4.3 and 4.4, and we obtain that

$$\Theta_{\infty}^{(\varepsilon)}(s) = \zeta(m+n)^{-1} \sum_{p, q \geq 1} \left( \int_{\mathbb{R}^{m+n}} f_{\varepsilon}(pa^s \bar{z}) f_{\varepsilon}(q \bar{z}) d\bar{z} + \int_{\mathbb{R}^{m+n}} f_{\varepsilon}(pa^s \bar{z}) f_{\varepsilon}(-q \bar{z}) d\bar{z} \right).$$

We also set

$$\begin{aligned}\Theta_{\infty}(s) &:= \zeta(m+n)^{-1} \sum_{p,q \geq 1} \left( \int_{\mathbb{R}^{m+n}} \chi(pa^s \bar{z}) \chi(q \bar{z}) d\bar{z} + \int_{\mathbb{R}^{m+n}} \chi(pa^s \bar{z}) \chi(-q \bar{z}) d\bar{z} \right) \\ &= 2\zeta(m+n)^{-1} \sum_{p,q \geq 1} \int_{\mathbb{R}^{m+n}} \chi(pa^s \bar{z}) \chi(q \bar{z}) d\bar{z}.\end{aligned}$$

We claim that

$$|\Theta_{\infty}^{(\varepsilon)}(s) - \Theta_{\infty}(s)| \ll \varepsilon^{1/2} e^{-(m+n-2)s/2}. \quad (6.17)$$

This reduces to the estimation of

$$\begin{aligned}& \left| \int_{\mathbb{R}^{m+n}} f_{\varepsilon}(pa^s \bar{z}) f_{\varepsilon}(q \bar{z}) d\bar{z} - \int_{\mathbb{R}^{m+n}} \chi(pa^s \bar{z}) \chi(q \bar{z}) d\bar{z} \right| \\ & \leq \left| \int_{\mathbb{R}^{m+n}} (f_{\varepsilon} - \chi)(pa^s \bar{z}) f_{\varepsilon}(q \bar{z}) d\bar{z} \right| + \left| \int_{\mathbb{R}^{m+n}} \chi(pa^s \bar{z}) (\chi - f_{\varepsilon})(q \bar{z}) d\bar{z} \right|. \quad (6.18)\end{aligned}$$

We observe that there exists  $0 < \nu_1 < \nu_2$  such that

$$\int_{\mathbb{R}^{m+n}} (f_{\varepsilon} - \chi)(pa^s \bar{z}) f_{\varepsilon}(q \bar{z}) d\bar{z} = 0$$

unless  $\nu_1 e^s q \leq p \leq \nu_2 e^s q$ , and by the Cauchy–Schwarz inequality under these restrictions,

$$\left| \int_{\mathbb{R}^{m+n}} (f_{\varepsilon} - \chi)(pa^s \bar{z}) f_{\varepsilon}(q \bar{z}) d\bar{z} \right| \leq \frac{\|f_{\varepsilon} - \chi\|_{L^2}}{p^{(m+n)/2}} \frac{\|f_{\varepsilon}\|_{L^2}}{q^{(m+n)/2}} \ll \frac{\varepsilon^{1/2}}{q^{m+n} e^{(m+n)s/2}}.$$

Since  $m+n \geq 3$ ,

$$\begin{aligned}& \sum_{p,q \geq 1} \left| \int_{\mathbb{R}^{m+n}} (f_{\varepsilon} - \chi)(pa^s \bar{z}) f_{\varepsilon}(q \bar{z}) d\bar{z} \right| \\ &= \sum_{q \geq 1} \sum_{\nu_1 e^s q \leq p \leq \nu_2 e^s q} \left| \int_{\mathbb{R}^{m+n}} (f_{\varepsilon} - \chi)(pa^s \bar{z}) f_{\varepsilon}(q \bar{z}) d\bar{z} \right| \\ &\ll \sum_{q \geq 1} \frac{\varepsilon^{1/2} e^s q}{q^{m+n} e^{(m+n)s/2}} \ll \varepsilon^{1/2} e^{-(m+n-2)s/2}.\end{aligned}$$

The sum of the other integral appearing in (6.18) is estimated similarly. This proves (6.17).

Provided that  $\varepsilon = \varepsilon(N) \rightarrow 0$ , the estimate (6.17) implies that

$$\left\| F_N^{(\varepsilon, L)} \right\|_{L^2(\mathcal{Y})}^2 = \Theta_{\infty}(0) + 2 \sum_{s=1}^{K-1} \Theta_{\infty}(s) + o(1).$$

Hence,

$$\left\| F_N^{(\varepsilon, L)} \right\|_{L^2(\mathcal{Y})}^2 \rightarrow \sigma^2 := \Theta_\infty(0) + 2 \sum_{s=1}^{\infty} \Theta_\infty(s)$$

as  $N \rightarrow \infty$ .

Finally, we compute the limit

$$\sigma^2 = \sum_{s=-\infty}^{\infty} \Theta_\infty(s) = 2\zeta(m+n)^{-1} \sum_{s=-\infty}^{\infty} \sum_{p, q \geq 1} \int_{\mathbb{R}^{m+n}} \chi(pa^s \bar{z}) \chi(q\bar{z}) d\bar{z}.$$

We note that the sum

$$\Xi := \sum_{s=-\infty}^{\infty} \chi \circ a^s$$

is equal to the characteristic function of the set

$$\{(\bar{x}, \bar{y}) \in \mathbb{R}^{m+n} : \|\bar{y}\| > 0, |x_i| < \vartheta_i \|\bar{y}\|^{-w_i}, i = 1, \dots, m\},$$

and

$$\begin{aligned} \int_{\mathbb{R}^{m+n}} \Xi(p\bar{z}) \chi(q\bar{z}) d\bar{z} &= \int_{1/q \leq \|\bar{y}\| < e/q} \left( \prod_{i=1}^m 2\vartheta_i \max(p, q)^{-1-w_i} \|\bar{y}\|^{-w_i} \right) d\bar{y} \\ &= 2^m \left( \prod_{i=1}^m \vartheta_i \right) \max(p, q)^{-m-n} \int_{1/q \leq \|\bar{y}\| < e/q} \|\bar{y}\|^{-n} d\bar{y} \\ &= 2^m \left( \prod_{i=1}^m \vartheta_i \right) \max(p, q)^{-m-n} \int_{S^{n-1}} \|\bar{z}\|^{-n} \left( \int_{q^{-1}\|\bar{z}\|^{-1}}^{eq^{-1}\|\bar{z}\|^{-1}} r^{-1} dr \right) d\bar{z} \\ &= 2^m \left( \prod_{i=1}^m \vartheta_i \right) \omega_n \max(p, q)^{-m-n}, \end{aligned}$$

where  $\omega_n := \int_{S^{n-1}} \|\bar{z}\|^{-n} d\bar{z}$ . We also see that

$$\begin{aligned} \sum_{p, q \geq 1} \max(p, q)^{-m-n} &= \sum_{p=1}^{\infty} p^{-m-n} + 2 \sum_{1 \leq p < q} q^{-m-n} \\ &= \zeta(m+n) + 2 \sum_{q \geq 1} \frac{q-1}{q^{m+n}} = 2\zeta(m+n-1) - \zeta(m+n), \end{aligned}$$

and thus

$$\sigma^2 = \sum_{s=-\infty}^{\infty} \Theta_{\infty}(s) = 2^{m+1} \left( \prod_{i=1}^m \vartheta_i \right) \omega_n \left( \frac{2\zeta(m+n-1)}{\zeta(m+n)} - 1 \right).$$

### 6.1 Proof of Theorem 6.1

As we already remarked above, it is sufficient to show that the average  $F_N^{(\varepsilon, L)}$  converge in distribution to the normal law. According to Proposition 3.4, it is sufficient to check that

$$\text{Cum}_{\mu_{\mathcal{Y}}}^{(r)} \left( F_N^{(\varepsilon, L)} \right) \rightarrow 0 \quad \text{as } N \rightarrow \infty$$

when  $r \geq 3$ , and

$$\left\| F_N^{(\varepsilon, L)} \right\|_{L^2(\mathcal{Y})} \rightarrow \sigma^2 \quad \text{as } N \rightarrow \infty.$$

These properties have been established above provided that the parameters

$$M = M(N), \quad \varepsilon = \varepsilon(N), \quad L = N^q, \quad \gamma = c_r(\log N), \quad K = K(N) \leq M(N)$$

satisfy the ten conditions (6.4)–(6.7), (6.10), (6.11), (6.13)–(6.16). It remains to show that such choice of parameters is possible. The condition (6.7) is guaranteed by the choice of  $L$ . First, we take

$$\varepsilon(N) = 1/N.$$

Then the first part of (6.5) holds. Then we select sufficiently large  $c_r$  in  $\gamma = c_r(\log N)$  so that (6.11) holds. After that we choose

$$K(N) = c_1(\log N)$$

with sufficiently large  $c_1 > 0$  so that (6.13) holds. Then it is clear that (6.16) also holds. Given these  $\varepsilon$ ,  $L$ ,  $\gamma$ , and  $K$ , we choose

$$M(N) = (\log N)(\log \log N)$$

so that the second part of (6.5), (6.6), (6.10), and (6.14) hold for all  $N \geq N_0(r)$ . With these choices, it is clear that (6.4) and (6.15) also hold. Hence, Theorem 6.1 follows from Proposition 3.4.

### 6.2 Proof of Theorem 1.2

For  $u \in M_{m,n}([0, 1])$ , we set



$$D_T(u) := \frac{\Delta_T(u) - C_{m,n} \log T}{(\log T)^{1/2}},$$

where  $C_{m,n} = 2^m \vartheta_1 \cdots \vartheta_m \omega_n$  with  $\omega_n := \int_{S^{n-1}} \|\bar{z}\|^{-n} d\bar{z}$ . We shall show that  $D_T(u)$  can be approximated by the averages  $F_N$  defined in (6.1). This will allow us deduce convergence in distribution for  $D_T$ . We observe that:

**Lemma 6.3**

$$\sum_{s=0}^{N-1} \int_{\mathcal{Y}} \hat{\chi}(a^s y) d\mu_{\mathcal{Y}}(y) = C_{m,n} N + O(1),$$

where  $C_{m,n}$  is defined above.

**Proof** We observe that

$$\sum_{s=0}^{N-1} \int_{\mathcal{Y}} \hat{\chi}(a^s y) d\mu_{\mathcal{Y}}(y) = \int_{\mathcal{Y}} \hat{\Xi}_N(y) d\mu_{\mathcal{Y}}(y),$$

where  $\Xi_N$  denotes the characteristic function of the set

$$\left\{ (\bar{x}, \bar{y}) \in \mathbb{R}^{m+n} : 1 \leq \|\bar{y}\| < e^N, \quad |x_i| < \vartheta_i \|\bar{y}\|^{-w_i}, \quad i = 1, \dots, m \right\}.$$

Using notation as in the proof of Proposition 4.6, we obtain

$$\begin{aligned} \int_{\mathcal{Y}} \hat{\Xi}_N(y) d\mu_{\mathcal{Y}}(y) &= \sum_{1 \leq \|\bar{q}\| < e^N} \sum_{\bar{p} \in \mathbb{Z}^m} \prod_{i=1}^m \int_{[0,1]^m} \chi_{\bar{q}}^{(i)}(p_i + \langle \bar{u}_i, \bar{q} \rangle) d\bar{u}_i \\ &= \sum_{1 \leq \|\bar{q}\| < e^N} \prod_{i=1}^m \left( \sum_{p_i \in \mathbb{Z}} \int_{[0,1]^m} \chi_{\bar{q}}^{(i)}(p_i + \langle \bar{u}_i, \bar{q} \rangle) d\bar{u}_i \right). \end{aligned}$$

We claim that

$$\sum_{p \in \mathbb{Z}} \int_{[0,1]^m} \chi_{\bar{q}}^{(i)}(p + \langle \bar{u}, \bar{q} \rangle) d\bar{u} = 2\vartheta_i \|\bar{q}\|^{-w_i}. \quad (6.19)$$

To prove this, let us consider more generally a bounded measurable functions  $\chi$  on  $\mathbb{R}$  with compact support, the function  $\psi(x) = \chi(x_1)$  on  $\mathbb{R}^m$ , and the function  $\psi(x) = \sum_{p \in \mathbb{Z}} \chi(p + x_1)$  on the torus  $\mathbb{R}^m / \mathbb{Z}^m$ . We suppose without loss of generality that  $q_1 \neq 0$  and consider a non-degenerate linear map

$$S : \mathbb{R}^m \rightarrow \mathbb{R}^m : \bar{u} \mapsto (\langle \bar{u}, \bar{q} \rangle, u_2, \dots, u_m)$$

which induced a linear epimorphism of the torus  $\mathbb{R}^m / \mathbb{Z}^m$ . Using that  $S$  preserves the Lebesgue probability measure  $\mu$  on  $\mathbb{R}^m / \mathbb{Z}^m$ , we deduce that

$$\begin{aligned} \sum_{p \in \mathbb{Z}} \int_{[0,1]^m} \chi(p + \langle \bar{u}, \bar{q} \rangle) d\bar{u} &= \int_{\mathbb{R}^m / \mathbb{Z}^m} \tilde{\psi}(Sx) d\mu(x) \\ &= \int_{\mathbb{R}^m / \mathbb{Z}^m} \tilde{\psi}(x) d\mu(x) = \int_{\mathbb{R}} \chi(x_1) dx_1, \end{aligned}$$

which yields (6.19).

In turn, (6.19) implies that

$$\int_{\mathcal{Y}} \hat{\mathbb{E}}_N d\mu_{\mathcal{Y}} = 2^m \left( \prod_{i=1}^m \vartheta_i \right) \sum_{1 \leq \|\bar{q}\| < e^N} \|\bar{q}\|^{-n}.$$

Using that  $\|\bar{y}_1\|^{-n} = \|\bar{y}_1\|^{-n} + O(\|\bar{y}_1\|^{-n-1})$  when  $\|\bar{y}_1 - \bar{y}_2\| \ll 1$ , we deduce that

$$\sum_{1 \leq \|\bar{q}\| < e^N} \|\bar{q}\|^{-n} = \int_{1 \leq \|\bar{y}\| < e^N} \|\bar{y}\|^{-n} d\bar{y} + O(1),$$

and expressing the integral in polar coordinates, we obtain

$$\int_{1 \leq \|\bar{y}\| < e^N} \|\bar{y}\|^{-n} d\bar{y} = \int_{S^{n-1}} \int_{\|\bar{z}\|^{-1}}^{\|\bar{z}\|^{-1} e^N} \|r\bar{z}\|^{-n} r^{n-1} dr d\bar{z} = \omega_n N + O(1).$$

This implies the lemma.  $\square$

Now we return to the proof of Theorem 1.2. Since

$$\Delta_{e^N}(u) = \sum_{s=0}^{N-1} \hat{\chi}(a^s \Lambda_u) + O(1),$$

Lemma 6.3 implies that

$$\|D_{e^N} - F_N\|_{C^0} \rightarrow 0 \quad \text{as } N \rightarrow \infty,$$

where  $(F_N)$  is defined as in (6.1). Therefore, it follows from Theorem 6.1 that for every  $\xi \in \mathbb{R}$ ,

$$|\{\xi \in M_{m,n}([0,1]) : D_{e^N}(u) < \xi\}| \rightarrow \text{Norm}_{\sigma}(\xi) \quad \text{as } N \rightarrow \infty.$$

Let us take  $N_T = \lfloor \log T \rfloor$ . Then

$$e^{N_T} \leq T < e^{N_T+1} \quad \text{and} \quad N_T \leq \log T < N_T + 1,$$

so that

$$D_T \leq \frac{\Delta_{e^{N_T+1}} - C_{m,n} N_T}{(\log T)^{1/2}} = a_T D_{e^{N_T+1}} + b_T$$

with  $a_T \rightarrow 1$  and  $b_T \rightarrow 0$  as  $T \rightarrow \infty$ . Hence, we deduce that

$$|\{u \in M_{m,n}([0, 1]) : D_T(u) < \xi\}| \geq |\{u \in M_{m,n}([0, 1]) : D_{e^{N_T+1}}(u) < (\xi - b_T)/a_T\}|.$$

It follows that for any  $\varepsilon > 0$  and sufficiently large  $T$ ,

$$|\{u \in M_{m,n}([0, 1]) : D_T(u) < \xi\}| \geq |\{u \in M_{m,n}([0, 1]) : D_{e^{N_T+1}}(u) < \xi - \varepsilon\}|.$$

Therefore,

$$\liminf_{T \rightarrow \infty} |\{u \in M_{m,n}([0, 1]) : D_T(u) < \xi\}| \geq \text{Norm}_\sigma(\xi - \varepsilon)$$

for all  $\varepsilon > 0$ . This implies that

$$\liminf_{T \rightarrow \infty} |\{u \in M([0, 1]) : D_T(u) < \xi\}| \geq \text{Norm}_\sigma(\xi).$$

A similar argument also implies the upper bound

$$\limsup_{T \rightarrow \infty} |\{u \in M_{m,n}([0, 1]) : D_T(u) < \xi\}| \leq \text{Norm}_\sigma(\xi).$$

This completes the proof of Theorem 1.2.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Björklund, M., Einsiedler, M., Gorodnik, A.: Quantitative multiple mixing. JEMS (**accepted**)
2. Björklund, M., Gorodnik, A.: Central limit theorem for group actions which are exponentially mixing of all orders. [arXiv:1706.09167](https://arxiv.org/abs/1706.09167)
3. Dolgopyat, D., Fayad, B., Vinogradov, I.: Central limit theorems for simultaneous Diophantine approximations. J. Éc. Polytech. Math. **4**, 1–36 (2017)
4. Eskin, A., Margulis, G., Mozes, S.: Upper bounds and asymptotics in a quantitative version of the Oppenheim conjecture. Ann. Math. **147**(1), 93–141 (1998)
5. Fréchet, M., Shohat, J.: A proof of the generalized second-limit theorem in the theory of probability. Trans. Am. Math. Soc. **33**(2), 533–543 (1931)
6. Fuchs, M.: On a problem of W. J. LeVeque concerning metric Diophantine approximation. Trans. Am. Math. Soc. **355**(5), 1787–1801 (2003)
7. Kleinbock, D.Y., Margulis, G.A.: Bounded orbits of nonquasiunipotent flows on homogeneous spaces. In: Sinai's Moscow Seminar on Dynamical Systems. American Mathematical Society Translations: Series 2, vol. 171. Advanced in Mathematical Sciences, vol. 28, pp. 141–172. American Mathematical Society, Providence (1996)
8. Kleinbock, D.Y., Margulis, G.A.: Flows on homogeneous spaces and Diophantine approximation on manifolds. Ann. Math. **148**(1), 339–360 (1998)

9. Kleinbock, D.Y., Margulis, G.A.: On effective equidistribution of expanding translates of certain orbits in the space of lattices. In: *Number Theory, Analysis and Geometry*, pp. 385–396. Springer, New York (2012)
10. Le Borgne, S.: Principes d'invariance pour les flots diagonaux sur  $SL(d, \mathbb{R})/SL(d, \mathbb{Z})$ . *Ann. Inst. H. Poincaré Probab. Stat.* **38**(4), 581–612 (2002)
11. Leveque, W.J.: On the frequency of small fractional parts in certain real sequences I. *Trans. Am. Math. Soc.* **87**, 237–260 (1958)
12. Leveque, W.J.: On the frequency of small fractional parts in certain real sequences II. *Trans. Am. Math. Soc.* **94**, 130–149 (1959)
13. Philipp, W.: Mixing sequences of random variables and probabilistic number theory. In: *Memoirs of the American Mathematical Society*, vol. 114. American Mathematical Society, Providence (1971)
14. Rogers, C.A.: Mean values over the space of lattices. *Acta Math.* **94**, 249–287 (1955)
15. Schmidt, W.M.: A metrical theorem in Diophantine approximation. *Can. J. Math.* **12**, 619–631 (1960)
16. Schmidt, W.M.: Asymptotic formulae for point lattices of bounded determinant and subspaces of bounded height. *Duke Math. J.* **35**, 327–339 (1968)
17. Siegel, C.L.: A mean value theorem in geometry of numbers. *Ann. Math.* **46**, 340–347 (1945)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.